



Università di Bologna

Dottorato di ricerca in
Metodologia Statistica per la Ricerca Scientifica
XIX ciclo

Program evaluation with continuous treatment:
theoretical considerations and
empirical application

Valentina Adorno

Dipartimento di Scienze Statistiche "P. Fortunati"
Marzo 2007



Università di Bologna

Dottorato di ricerca in
Metodologia Statistica per la Ricerca Scientifica
XIX ciclo

Program evaluation with continuous treatment:
theoretical considerations and
empirical application

Valentina Adorno

coordinatore

prof. Daniela Cocchi

tutor

prof. Attilio Gardini

co-tutor

prof. Guido Pellegrini

Settore Disciplinare
SECS-S/03

Dipartimento di Scienze Statistiche “P. Fortunati”
Marzo 2007

Preface

The ambitious aim of the thesis is to develop a matching estimator approach to evaluate causal effects of a policy intervention on some outcome variables in a continuous treatment framework.

The main motivation of the research project derives from a raising and strong demand of modern welfare states for objective knowledge about the effects of various government interventions and activities; someone can benefit and other can lose from such programs. Assessment of these benefits and losses often plays an important role in policy decision-making. In particular, the evaluation problem of concern here is the ex-post measurement of the impact of a policy reform or intervention on a set of well-defined outcome variables.

Most of the relevant literature on the program evaluation deals with the estimation of causal effects of a binary treatment on one or more outcomes of interest in a non-experimental framework. In practice, however, treatment regimes need not to be binary and individuals might be exposed to different levels or doses of treatment. In these situations, studying the impact of such treatment as if it were binary can mask some important features on it. Moreover other parameters might be of interest. It could be interesting, for example, learning about the form of the entire function of average treatment effects over all possible values of the treatment levels.

The idea of estimating causal effects of a public intervention when the treatment variable is continuous comes from an economic problem we want to solve. It refers to the evaluation of the impacts of the Law 488/92 in Italy, that is the most important policy intervention to subsidize private capital accumulation in the poorest Italian regions in the last decade. The questions we want to answer mainly refer to two aspects. Does the law 488 affect performances of subsidized firms? And, are the effects different with respect to the amount of public subsidies? Then, the idea is to use the amount of received subsidy as a continuous treatment variable in order to estimate causal effects as a function of the different treatment levels.

The empirical application has some particular characteristics that suggest to use a matching estimation approach among the other methods commonly adopted in the program evaluation literature. Then, some matching estimators will be proposed, together with the assumptions they require in order to estimate causal effects.

Then, the main objectives of this thesis refer to an analysis of program evaluation in a continuous treatment setting, the development of an appropriate matching estimation method and the analysis of impact of differences in treatment level on policy outcomes.

As empirical application, is proposed the case of the Law 488/1992 for the manufacturing sector in Italy. The final results show that the impact of L. 488 on subsidized firms is positive and statistically significant: the firm growth outcome variables increase higher in the subsidized firms than in non subsidized ones. Furthermore we find that higher the level of incentive, higher the policy effect until a certain point, from which the marginal impact is decreasing.

Analysis have been obtained using the statistical softwares STATA and SAS.

Structure of the thesis

An introduction to the program evaluation problem in the economic statistic context together with some considerations on the interesting parameter to estimate is contained in the Chapter 1. The aim is to contextualize in a statistical vision the problems to solve in order to estimate causal effects of an intervention.

Chapter 2 concerns statistical tools used in program evaluation with a binary treatment regime context. The aim is to give an idea to the reader about the state of the literature in this field, and at the same time to explain methods that are important building blocks for the further generalization to more complicate settings, such as a continuous treatment regime, that is introduced in Chapter 3. It refers to an overview on the relevant literature of programme evaluation with continuous treatment. From this, some considerations on the proposed approaches are discussed. They constitute the starting point of our research project (Chapter 4). A matching estimation approach is proposed in order to estimate causal effects at particular values of the treatment levels. Chapter 5 describes the dataset used in the application, while Chapter 6 contains the final results of the application. Some final comments and conclusions are contained in the Chapter 7, with a brief introduction to further developments.

Acknowledgements

This thesis is the result of my three-year period of research at the Statistics Department of the University of Bologna.

It is my great pleasure to thank all people who helped me in some ways in the realization of this work. I would like to thank Prof. Guido Pellegrini for his supervision, precious suggestions and inspiration. He introduced and involved me in this interesting research project, giving me the chance to learn and grow really much, not only from a professional point of view. I am particularly grateful to Cristina Bernini for her suggestions and constant and encouraging support. I wish to thank Prof. Attilio Gardini for the realization of this thesis.

My period of study at the University College of London was crucial for my research, so I would like to thank all the people who made it possible. A special thank to Alessia Matano for being a kind roommate and still a special friend. I am very grateful to my PhD colleagues, Mariagiulia Matteucci and Laura Sardonini, who always unconditionally helped me. They are not only precious colleagues but very special friends. I also would like to thank Marta Disegna, Caterina Liberati and Giulia Roli for the fun time and talks we had together.

Many thanks to all people I had close to me in these years. A very special thank goes to my family, who always trusts and supports me, and to my friend Alex for his patience. A particular thought to Giulia who always helps me to keep in mind how much important is a smile. Last but not least I would like to thank Furio for being close to me in this final period: most of my energy comes from him.

Contents

Preface	i
1 Program evaluation	9
1.1 Introduction	9
1.2 The Evaluation Problem	9
1.3 Estimation of the counterfactual: problems to solve	11
1.4 Which Parameter of Interest?	12
2 The binary treatment case	15
2.1 Introduction	15
2.2 The Parameters of Interest	16
2.2.1 Homogeneous Treatment Effects	18
2.2.2 Heterogeneous Treatment Effects	19
2.3 Experimental Data	20
2.3.1 Experimental Data in Practice	22
2.4 Non-Experimental Data	23
2.5 Instrumental Variable Estimator	25
2.5.1 Homogeneous Treatment Effect	26
2.5.2 Heterogeneous Treatment Effect	27
2.5.3 Instrumental variable approach in practice	30
2.6 Heckman Selection model	32
2.6.1 Homogeneous Treatment Effect	32
2.6.2 Heterogeneous Treatment Effect	34
2.6.3 Selection models in practice	35
2.7 Difference-in-Difference Estimators	35
2.7.1 Trend Adjusted Diff-in-Diff	37
2.7.2 Difference in difference in practice	38
2.8 Matching Estimators	39

2.8.1	The Propensity score	45
2.8.2	Matching Diff-in-Diff Approach	47
2.8.3	Matching approach in practice	48
2.9	Regression Discontinuity Estimators	49
2.9.1	Regression discontinuity Design in practice	53
3	The continuous treatment case	55
3.1	Introduction	55
3.2	From binary to continuous treatment	55
3.3	An overview on the literature	57
3.4	Generalized propensity score: parametric approach	58
3.5	Some non-parametric approaches	61
3.5.1	Subclassification on the propensity score	63
3.5.2	Matching estimators	65
3.6	Conclusions: our starting point	71
4	A new approach to empirical estimation	73
4.1	Introduction	73
4.2	The continuous treatment setting	73
4.2.1	The selection process	75
4.3	The parameters of interest	77
4.3.1	Average treatment level effects	78
4.3.2	Treatment dose function	79
4.4	The random treatment level case	81
4.4.1	Estimation of the effects: a matching approach	82
4.4.2	Test for the independence assumption	85
4.5	The non-random treatment level case	87
4.6	A new approach: the use of a matching procedure	89
4.6.1	Structured form of the selection process	90
4.6.2	Non-structured form of the selection process	96
4.7	Reduction of the multidimensionality	98
4.7.1	Structured selection process	99
4.7.2	Non-structured selection process	101
5	The law 488/1992 in Italy	103
5.1	Introduction	103
5.2	Capital subsidies in Italy	103

5.3	The Law 488/92	106
5.3.1	The higher applicable subsidy	107
5.3.2	Why not a RDD approach?	109
5.4	The data implementation	110
6	Application	115
6.1	Introduction	115
6.2	The treatment variable: amount of subsidies	115
6.3	The outcome variables	117
6.4	Structured from of the selection process	118
6.4.1	Impact of the Law 488	120
6.5	Non-structured from of the selection process	130
6.5.1	Impact of the Law 488	130
7	Conclusions	139
	Bibliography	141
	Appendices	149
A	Integration of datasets	151
B	Outcome variable distributions	155

List of Tables

5.1	Intensity of the subsidies	108
5.2	Distribution of projects according to main characteristics (1996-2000) . .	113
5.3	Summary of main covariates in the final dataset (1996-2000)	114
6.1	Summary of the granted subsidies	116
6.2	Summary of the percent share of subsidies on the total investment . . .	116
6.3	Summary of outcome variables	118
6.4	Structured form: Propensity score estimate	119
6.5	Structured form: Level of treatment estimate	120
6.6	Structured form: TTE estimates radius=fix	121
6.7	Structured form: TTE estimates radius=mean	121
6.8	Structured form: TTE estimates radius=std	122
6.9	Structured form: TTE estimates radius=mean	122
6.10	OLS estimates: impact of the amount of subsidies on the treatment level (structured case)	124
6.11	Quantile estimates: Employment (structured case)	126
6.12	Quantile estimates: Turnover and fixed assets (structured case)	127
6.13	Non structured form: Propensity score estimate	130
6.14	Non-structured form: TTE estimates radius=fix	131
6.15	Non-structured form: TTE estimates radius=mean	131
6.16	Non-structured form: TTE estimates radius=std	131
6.17	Non-structured form: TTE estimates radius=mean	132
6.18	OLS estimates: impact of the amount of subsidies on the treatment level (non structured case)	133
6.19	Quantile estimates: Employment (non structured case)	135
6.20	Quantile estimates: Turnover and Fixed assets (non structured case) . .	136

A.1	From the original to the matched dataset	152
A.2	Impact of data matching process on eligible projects distribution	154

List of Figures

6.1	Distribution of the treatment variable	117
6.2	Predicted OLS estimates for treatment impact on the amount of subsidy (structured case)	125
6.3	Kernel estimates for treatment impact on the amount of subsidy (struc- tured case)	128
6.4	Kernel estimates for treatment impact on the amount of subsidy (struc- tured case)	129
6.5	Predicted OLS estimates for treatment impact on the amount of subsidy (non structured case)	134
6.6	Kernel estimates for treatment impact on the amount of subsidy (non structured case)	137
6.7	Kernel estimates for treatment impact on the amount of subsidy (non structured case)	138
B.1	Distribution of the outcome variable: turnover	155
B.2	Distribution of the outcome variable: employment	156
B.3	Distribution of the outcome variable: fixed assets	156
B.4	Distribution of the outcome variable: gr. margin on turnover	157
B.5	Distribution of the outcome variable: per capita turnover	157
B.6	Distribution of the outcome variable: debt charges on debt stock	158

Chapter 1

Program evaluation

1.1 Introduction

A common characteristic of the modern welfare state is the demand for objective knowledge about the effects of various government interventions and activities; someone can benefit and other can lose from such programs. Assessment of this benefits and loses often plays an important role in policy decision-making. In this introduction, we want to provide an understanding overview of the evaluation problem and to outline how, along with adequate economic and statistical analysis, program evaluation can play a fundamental role in modern welfare states.

1.2 The Evaluation Problem

Recently the literature on the economic statistic concerning the role of economic policies and their ability to produce effects in one desired direction is vast and continues to growth, as many economics with modern welfare states have floundered and costs of running welfare states have been increased. Even if the importance of this kind of matters is well acknowledged, it refers to a complex argument, which is difficult to completely define. In general, a clearly and complete definition of the program evaluation concept doesn't exist, because of its different fields of application and aims at which is referred. It can include very different concepts depending on the idea of the various authors.

The evaluation problem of concern here is the measurement of the impact of a policy reform or intervention on a set of well-defined outcome variables. Some example of these

policies can be childcare subsidy, training program and business incentives; relative outcome variables could be, respectively, child's exams results, individual employment earnings or duration, job creation or business earnings. There are many references in the literature which document the development of the evaluation policy analysis in economics. In the labour market some examples are Heckman and Robb (1985), Heckman *et al.* (1999) and Blundell *et al.* (2001) and Ichino *et al.* (2003) for Italy. Some examples of application of economic development program evaluation aimed at influencing business are Bondonio (2000) and Bondonio (2004), Bartik and Bingham (1995), Boarnet and Bogart (1996), Dowall (1996) and Carlucci and Pellegrini (2003) for Italy.

In general, these kind of studies could be classified by the level of analysis: macroeconomics evaluation is aimed to quantify effects on territory and economic system together with spillover implications, while in microeconomic evaluation the attention is on the individual effects of the units which the program is addressed to. However, in both cases, the main interest is the causal relationship between the program and the observed variations in the variables of interest, that is the causal effect of a "treatment" on an "outcome". It should be noted that this is not the only potential evaluation question, but its importance is certainly fundamental to understand what kind of result has been obtained.

At this point, the main problem is to identify what is, and how can be quantified an *effect*; the first step for a correct evaluation question is to clearly define *on what* effect we are looking for. This could be done specifying one or more variables of interest (outcomes), that identify the observable features of the population the policy is addressed to. Secondly, it is fundamental to know the effect *of what* we are looking for; it is necessary to identify one variable that denote the treatment state. In this sense, the concept of effect denotes the variation caused by a policy intervention, and the evaluation analysis concerns the methods applied to empirically verify the causality of the intervention. This concept of causality plays a critical role in program evaluation. Some references about this can be founded in Ichino (2002).

The evaluation problem, therefore, is to measure the impact of the program on an outcome of interest (Y) for each individual; this is commonly defined as the difference between the value of Y observed after the units have been exposed to the intervention, and the value that Y would have taken if the same unit had not being exposed to the treatment. This latter value is defined as the *counterfactual*. This framework follows the potential outcome approach as described and developed by Rubin (1974); it means that ex-post only the outcome corresponding to the program in which the individuals

participate is observed. That is why it is called evaluation problem: it can be regarded as a missing-data problem since, at a moment in time, each unit is either in the program under consideration or not, but not both. If we could observe the outcome variable for those in the program had they not participated, there would be no evaluation problem.

1.3 Estimation of the counterfactual: problems to solve

To solve the evaluation problem described above, the focus of all approaches presented in literature is on the estimation of missing counterfactual data; they mainly differ in the assumptions they make about how the missing data are related to the available information. In any case, there are two major problems that arise in estimation of the counterfactual: first, there could be some factors, independent from the implementation of the program being evaluated, that can influence the outcome variable Y . The second problem refers to the non-random nature of the selection process that designate treated individuals.

The first problem derives from the fact that changes in the general conditions of treated units may occur during the program intervention for reason independently from the latter. For example, these types of variation may be caused by a general economic growth of target individuals, and one may be mistakenly induced to overestimate the real impact of the program. This type of problem is called *omitted variable bias*. It is typical of one group design evaluation strategies, that is when only data from treated units are considered. In this case the outcome variable is usually measured at different times, pre and post intervention: mean impact estimations are obtained as pre-post intervention differences. Nevertheless, they are unbiased estimations of the effect only if there are no differences between the pre-intervention values of Y of the untreated and the outcome values they had obtained if they were treated (counterfactual). If there is some difference, it represents the degree of omitted variable bias: such a variation could arise when there is an impact on the outcome due to some economic or social factors, uncorrelated to the program intervention, that would have increased, or decreased the value of Y even in the absence of the program. As a result, the main problem that an evaluator has to deal with when a one group design strategy is adopted, is to control for any exogenous factors that may affect the outcome during the program intervention.

The second difficulty in program evaluation arises because treatment assignment is typically based on pre-intervention economic and social conditions. Nevertheless, at the beginning of the program, target units may have a significantly different set of

conditions from the excluded individuals. It is typically the case when impact estimates are obtained using a comparison group design strategy. A source of bias can be arisen because the effect estimates are obtained by comparing the output of the treated and the excluded; these are unbiased only if the difference between the pre-intervention set of condition of the two groups is equal to zero. This type of variation measures the size of the bias, referred to as *selection bias*, affecting the impact estimates. Hence, when dealing with this strategy it is crucial to properly control for the intervention differences, between treated and untreated units.

1.4 Which Parameter of Interest?

Another important issue that has to be taken into consideration to better understand and fully describe the program evaluation problem regards the parameters and the quantities that might be useful in the analysis of a modern welfare economy.

There are many possible counterfactuals of interest for evaluating a program; for example it could be interesting to compare the state of the world in the presence of the intervention to the state of the world if the program did not exist at all, or if alternative programs were used. A full evaluation should consider all outcomes of interest for all units, both in the current state and in all alternative state of interest, and a mechanism for valuing these outcomes in the different states.

The fundamental elements of a well-done evaluation research, once appropriate outcomes measures has been identified, include the direct benefits received, the level of behavioral variables and the payment for the program for both participants and non participants.

The conventional econometric and statistic evaluation literature usually defines the effect of participation to be the *direct effect* of the program on participants explicitly enrolled in the program. It ignores the effects of a program that do not flow direct participation, known as the *indirect effects*, and equates “treatment” outcomes with the direct outcome variable in the program state and “no treatment” with the direct outcome variable in the no program state. Among possible outcomes of interest, the traditional econometric literature focuses on counterfactual means, with the conventional assumption that the no treatment state approximates the no program state; this would be true if indirect effects are negligible. The transition from the individual to the group level counterfactual recognizes the impossibility of observing the same person in both states at the same time. This represents the statistical solution to the problem of causality:

given that the causal effect for a single individual i cannot be observed, the statistical solution proposes to compute the average causal effect for the entire population or for some interesting sub-groups.

This wants to be just a brief introduction of the issue regarding the interesting quantities to estimate. As it might be easily understood, it represents the starting point of any proposed studies in the evaluation literature and it depends on the framework the analysis refer to. In particular, there is a fundamental distinction that will characterize this thesis: the program evaluation issue will be studied starting with the easier case of a binary treatment intervention. The next step is to extend the analysis and to widen its applicability by allowing for continuous treatment regimes. Thus, in the next chapters the issue regarding the interesting quantities to estimate will be better defined, and the traditional parameters of interest will be presented and discussed, keeping separated this two treatment regime cases.

Chapter 2

The binary treatment case

2.1 Introduction

This chapter concerns statistical tools used in program evaluation with a binary treatment regime context. The attention is focused on statistical techniques for solving the evaluation problem in order to estimate impacts of an intervention on some well defined outcomes of interest. The aim is to give an idea to the reader about the state of the literature in this field, and at the same time to explain methods that are important building blocks for the further generalization to more complicate settings, such as a continuous treatment regime.

The binary treatment case represents the starting point not only for our analysis, but for all studies in the program evaluation literature. It represents the most intuitive and easier case, and in some situation the only possible one. It is the case, for example, of an equal medical treatment for some patients or a job training program for some individuals. In this situation, units can be classified only in two group: the participants and the non-participants.

In the following sections, after a formal description of the binary treatment setting and of the conventional parameters of interest, there will be given an overview of the statistical tools and traditional approaches proposed in literature to estimate the effects of a program, following the paper of Blundell and Costa Dias (2002), that summarizes the alternative approaches to evaluation in empirical microeconometrics.

2.2 The Parameters of Interest

As mentioned in the previous chapter, there are many possible counterfactuals of interest for evaluating a program. Once established that the objective of the analysis is the ex-post evaluation of impacts of an intervention, the statistical solution to the problem of causality is represented by the transition from the individual to the group level counterfactual: given that the causal effect for a single individual cannot be observed, the solution is to compute the average causal effect for the entire population or for some interesting sub-groups. Following this part of literature and using the now standard potential outcome notation, introduced by Rubin (1974), we can consider N units, indexed by $i = 1, \dots, N$, viewed as drawn randomly from a large population, and a policy intervention introduced at time k , for which we want to measure the impact on some outcome variable Y . Let consider the case where D is a dummy variable representing the treatment status, assuming value 0 if the unit has not been treated and 1 otherwise. With this binary treatment variable, each units is characterized by a pair of potential outcome: Y_{it}^1 for the outcome under the active treatment, Y_{it}^0 otherwise. This outcome Y is assumed to depend on a set of covariates X through a particular relationship concerning the participation status in each period t . The equation of the outcomes can be generically represented as:

$$\begin{aligned} Y_{it}^1 &= f_t^1(X_i) + U_{it}^1 \\ Y_{it}^0 &= f_t^0(X_i) + U_{it}^0 \end{aligned} \tag{2.1}$$

where the subscripts i and t identify the unit and the time period, while the superscript stands for the treatment status. The functions f^0 and f^1 represent the relationships between the set of covariates X and the potential outcome Y^0 e Y^1 , while the terms U^0 e U^1 identify the mean zero error terms, assumed to be uncorrelated with the X . This vector of variables is assumed known at the participation decision time, and not influenced by treatment. That is why the time subscription is excluded. Then, the general observed outcome Y can be written as:

$$Y_{it} = D_{it}Y_{it}^1 + (1 - D_{it})Y_{it}^0. \tag{2.2}$$

When $D = 1$ we observed Y^1 , when $D = 0$ we observed Y^0 .

The definition of potential outcome already made implicitly uses the assumption of no interference between units or, the so called, stable-unit-treatment-value assumption

(SUTVA). It assumes that the potential outcome Y^0 and Y^1 of individual i are not affected by the allocation of other individuals to the treatments. In other words, it is assumed that the observed outcome Y_{it} depends only on the treatment to which individual i is assigned and not on the allocation of other individuals.

Moreover we assume that the participation decision can be identify by the following rules:

$$I_i = h(W_i) + V_i$$

that means for each units there is an index I , function of the set of variable W , for which participation occurs when it raises above zero. V_i is the error term and

$$\begin{aligned} D_{it} &= 1 \quad \text{if } I_i > 0 \quad \text{and } t > k \\ D_{it} &= 0 \quad \text{otherwise.} \end{aligned}$$

When specifical application are considered, there are several fundamental decision to be taken in account: first, except for experimental data, participation to treatment is not random. For that reason the correlation between the outcome error term (U^0, U^1) and the enrolment in the program, represented by D_{it} , could not be equal to zero. That could be happen because individual non-observable characteristics leading to the participation decision could affect the outcome Y , and thus some correlation between the error term and the participation variable is expected. Any method that fails to consider this problem is not able to identify the correct parameter of interest. Secondly, an important decision to take, is whether to assume heterogeneous or homogeneous treatment effects. Typically, we do not expect the policy intervention affects all individual in exactly the same way. That is, there will be heterogeneity in the impact across units. Consequently, possible questions that evaluation methods attempt to answer could be different. As previous mentioned, the most commonly in the evaluation literature are average effects on individuals of a certain type. To be more precisely, the different parameter of interest measured in period $t > k$ can be expressed as:

- **Average Treatment Effect:**

$$\alpha_{ATE} = E[\alpha_{it} | X = X_i]$$

that represents the population average treatment effect,

- **Average Treatment on the Treated Effect:**

$$\alpha_{TTE} = E[\alpha_{it}|X = X_i, D_t = 1]$$

that is the effect of treatment on units that were assigned to treatment,

- **Average Treatment on the Untreated Effect:**

$$\alpha_{TU} = E[\alpha_{it}|X = X_i, D_t = 0]$$

that is commonly interesting for decision about some treatment extended to a group formally excluded from the treatment.

The parameter that had received more interest in the current literature is the α_{TTE} ; Heckman and Robb (1985) and Heckman *et al.* (1997b) argued that the subpopulation of treated units is often of more interest than the overall population in the context of narrowly targeted programs. For example, if a program is to encourage business in particular depressed areas, there is often little interest to evaluate such program on areas that has not this kind of disadvantages. Of course the treatment D might have an effect not only on the mean of the outcome, it might influence the whole distribution. In some situation this could be what we are interested in, in other could not be.

Only under the assumption of homogeneous treatment effect all these parameter are identical, but it is obviously not true when treatment effects vary across individuals. Let consider separately this two different situations.

2.2.1 Homogeneous Treatment Effects

This is the simplest case; effect is assumed to be constant across units. So that

$$\alpha_t = \alpha_{it}(X) = f_t^1(X_i) - f_t^0(X_i) \quad \text{with } t > k, \forall i$$

This means that all people gain or lose the same amount in going from the state “0” to the state “1”. So long as $U^0 = U^1$ among people with the same X , there is no heterogeneity in the outcomes moving from “no treated” versus “treated”. The outcome equation (2.1) can be re-written as

$$Y_{it} = f_t^0(X_i) + \alpha_t D_{it} + U_{it} \quad (2.3)$$

However this assumption of absence of heterogeneity in response to treatment is very strong, and when tested, it is almost always rejected (see Heckman *et al.* (1997a)).

2.2.2 Heterogeneous Treatment Effects

It seems reasonable to assume that treatment impact varies across individuals; this could be come systematically through the observable component, or be a part of the unobservables. The outcome equation (2.2) can be written as:

$$Y_{it} = D_{it}Y_{it}^1 + (1 - D_{it})Y_{it}^0 = f_t^0(X_i) + \alpha_t(X_i)D_{it} + [U_{it}^0 + D_{it}(U_{it}^1 - U_{it}^0)] \quad (2.4)$$

where

$$\alpha_t(X_i) = E[\alpha_{it}(X_i)] = f_t^1(X_i) - f_t^0(X_i)$$

is the expected treatment effect at time t among units characterized by X_i .

Obviously, the additional problems with this heterogeneous setting concern the observables and their role in the identification of the parameter of interest and the form of the error term. This can differ across observations, according to the treatment status of each unit. If there is selection on observables, the OLS estimator, after controlling for the covariates X , is inconsistent for $\alpha_t(X)$, identifying the following parameter:

$$E[\hat{\alpha}_t(X)] = \alpha_t(X) + E[U_t^1|X, D_{it} = 1] - E[U_t^0|X, D_{it} = 0] \quad \text{with } t > k$$

There is one case when, even if there is a dispersion in the treatment effect and the unobservables components are different among the two groups ($U_{it}^1 \neq U_{it}^0$), the two parameters of interest α_{ATE} and α_{TTE} are still equal. This is the case when

$$E[U_{it}^1 - U_{it}^0|X = X_i, D_{it} = 1] = 0$$

that arises when, conditional on X , D does not explain or predict $U_{it}^1 - U_{it}^0$. This is the case when individuals who select into state 1 or 0 do not know $U_{it}^1 - U_{it}^0$ in making their decision to participate in the program.

The distinction between a model with $U^1 = U^0$ and one with $U^1 \neq U^0$ is fundamental to understanding modern developments in the program evaluation literature. When we condition on X and $U^1 = U^0$, it means that everyone with the same value of X has the same treatment effect, and this is a quite strong assumption. Nevertheless, this setting greatly simplifies the evaluation problem; one parameter is able to answer all of the conceptual evaluation question we have presented. The “treatment on the treated effect” can be viewed as the same effect of taking an individual at random and putting him into the program. Otherwise, when $U^1 \neq U^0$ there are a variety of different treatment effects that can be defined. In this case, conventional econometric procedures often break down or require substantial modifications.

2.3 Experimental Data

One of the solution of the evaluation problem could be the experimental data setting because, as mentioned previously, it provides the correct missing counterfactual. In recent years, the use of experimental designs, for example to evaluate North America employment and training programs, has rapidly increased. This approach has been less common in Europe, though a small number of experiments have been conducted in Britain, Norway and Sweden. The impact estimate of these experiments, often called *social experiments*, are easy for analyst to calculate and for policymakers to understand. They also can be used to study the properties of the other methodologies applied to program evaluation. That is, a comparison of results from non experimental data to those obtained from experimental data can help to assess appropriate methods where experimental data are not available.

In many ways this method is the most convincing one, since it directly constructs a comparison, or control, group. In this way it may be overcome the missing data problem and avoided the problem of causal effects identification. The advantages of experimental data are discussed in Hausman and Wise (1985). They were based on earlier statistical experiment developments (see for example Fisher (1951)).

The contribution of this type of data is to rule out bias caused by self-selection, as individuals are randomly assigned to the program. By design, the experiment D will be independent of any kind of influence on the outcome Y , whether observed or unobserved. To see why, let image an experiment in which is chosen to participate in a program a random sample from a group of eligible units. Within this focus group, assignment to treatment is completely independent either from any outcome or the

treatment effect. Under the assumption of no spillover effect, the two groups, treated and excluded, are statistically equivalent in all dimension, except treatment status.

In the case of homogeneous effect the treatment impact can be measured by a simple difference between the mean outcomes of the treated and the untreated:

$$\hat{\alpha}_{ATE} = \hat{\alpha}_{TTE} = \bar{Y}_t^1 - \bar{Y}_t^0 = E[Y_{it}|D_{it} = 1] - E[Y_{it}|D_{it} = 0] \quad t > k \quad (2.5)$$

where \bar{Y}_t^1 and \bar{Y}_t^0 are respectively the mean of the treated and non-treated outcomes at time t after the program. Notice that, within the context of the model of equation (2.1), the method of social experiment does not set either $E[U_{it}^1|X, D = 1]$ or $E[U_{it}^0|X, D = 1]$ equal to zero. It simply balances the selection bias in the treatment and control groups.

To obtain an estimator of the causal effect in (2.5) it might be simply taken the difference in the sample mean equivalent of the population moments between the two groups. There is no particular problem to calculate the causal effects in this way, but it can be useful to derive it via regression model. Let consider the model

$$Y_{it} = \beta_{0t} + \beta_{1t}D_{it} + \varepsilon_{it} \quad \text{with } t > k.$$

The OLS estimate of β_{1t} will be an estimation of the causal effect since is the difference in the mean of the outcome Y between the control and the treatment group. The regression model presented here has the advantage that it will also give immediately an estimation of the standard error on the coefficient of D that can be used for inference purposes. Another advantage derives from the fact that it can naturally be generalized to the case of continuous treatment simply running a regression of Y on D . Obviously the assumption of the linearity relationship between the two variables is quite strong and must be verified.

As previous mentioned, when we move into the case of heterogeneous effects we need to provide some summarize of the treatment impact because of the impossibility of observing the same person in both states at the same time. This raises the question of what kind of effect we are interested in. In the homogeneous structure this is not a meaningful question, because all the parameter of interest mentioned above are equal. It is easy to demonstrate that the average treatment effect, ATE, can be consistently estimated by the coefficient on D_{it} of the regression of Y_{it} on D_{it} (see Stock and Watson

(2003)). Simply re-write the model in this way:

$$Y_{it} = \beta_{0t} + \bar{\beta}_{1t}D_{it} + (\beta_{1it} - \bar{\beta}_{1t})D_{it} + \varepsilon_{it} = \beta_{0t} + \bar{\beta}_{1t}D_{it} + u_{it}$$

Note that the unobservable component u_{it} is mean-independent of D_{it} so that the OLS estimate of the coefficient on D will be the ATE.

Of course social experiments have their own drawbacks: they might not be feasible at all for political, ethical, logistic, or financial reason and they are rare in economics and typically expensive to implement. Moreover, the result of this type of estimation can be invalidated by a number of disrupting factors associated with the experimental design: for example there could be some dropouts, especially among the non-treatment group (see Heckman *et al.* (1999)). This process could be non-random and the statistical equivalence should not more hold. To ensure random assignment, at least with respect to observables, it can be compared the observable characteristics of both the treatment and control groups. Moreover, further differentiating factors are introduced when some experimental controls may search for alternative programs and likely to succeed. On the other hand, some treatment units can decide to leave the program. This is the so called problem of the *partial compliance*. Furthermore, the individual may also change their behavior as a consequence of the experiment itself. All these kind of factors may invalidate the fundamental characteristic of the experimental approach and the consistency of the presented estimator. Finally, most practical applications suffer the fact that samples are small. This usually implies the presence of observable factors X , which are not be balanced completely by the process of randomization.

2.3.1 Experimental Data in Practice

A detailed list of the different social experiments taken place during the 60th, especially in the US, can be found in Greenberg and Shroder (1997). Among these several social experiments that have been carried out, the best known are the training and employment experiments of the United States: the National Supported Work Demonstration (NSWD) and the National JTPA Studies.

The NSWD, one of the first training and employment social experiment, was carried out in 10 different sites across the US. It was planned to help workers in disadvantaged conditions, in particular women in receipt of AFDC (aid for families with dependent children), ex-criminal transgressor, ex-drugs addicts, and in general economically disadvantaged youths. The program provided the applicants with a guarantee job for 9 to

18 months. Qualified candidates were randomly assigned to treatment: one half of the individuals were assigned to the treatment group and the other half was precluded to receive it and was assigned to the control group. The impact of the intervention was measured on the earnings and on rate of employment.

This program is an example of the so called *pilot programs* or *demonstrations*, in which the intervention are quite easy to implement and conduct. That is why they are stronger from the “internal validity” point of view. On the other hand, it is more difficult to generalize outside the conclusions drawn from this kind of experimental situation because for example the sample and the program might be non representative. In general, they have weak properties in term of “external validity”.

It is opposite the case of the so called *ongoing programs*, in which the results are more robust but they are very difficult and complicated to implement. This is the case of the JTPA studies carried out between the 1984 and 1996 in 16 sites of the US (see Orr *et al.* (1996)): the focus of the analysis is on the evaluation of different ongoing training and employment randomly assigned programs on the social-economic situation of disadvantaged individuals.

Another application of a ongoing social experiment is the study of the Self-Sufficiency Project (see Card and Robins (1998)): it evaluates labor supply responses of single mother in British Columbia. Half of the candidate were recorded, but they were randomly executed from the program. The analysis shows a significant evidence of the effectiveness of financial incentives in inducing welfare receipt into work.

Social experiment are also useful to understand and to study the properties of other methodologies applied to program evaluation. For example the main focus of the studies carried out by LaLonde (1986) is to underline the difference between experimental and non-experimental data. This has been done by comparing the results obtained applying different type of non-experimental estimation methodologies using the same experimental dataset.

2.4 Non-Experimental Data

In program evaluation the randomization assignment is generally considered as the gold standard (Orr (1999)). In the natural sciences, in particular in medical application, this is especially true; interventions are randomized to different people on an individual basis. Policy interventions affecting in general the economic system are certainly different; often they do not lend themselves to control for implementation, and even more often

they are implemented before a controlled experiment can be designed and executed.

By contrast to the experimental analysis, in non-experimental, or observational, studies the data are not derived in a process that is completely under the researcher control. For example a government authority might have offered the program to particular areas or specific individuals, in order to improve their condition or because it is believed that they held favorable expectations regarding the impact of the program.

The main objective of any observational study is to use observable information in an appropriate way, in order to replace the comparability of treatment and control group by an appropriate alternative identification condition. Analysts must replace the counterfactual missing data with data on the non-participants, together with some assumptions that can make comparable the two groups. The objective is to use the available information such that, in sub-population defined by these observables any remaining differences between treated and non-treated might be attributed to the program. Hence, in general, non-experimental data are more difficult to deal with and require special care. Another problem arises if it is considered that in general there exists a set of assumptions necessary to identify the effect of a program that cannot be verified with the data. It also happens with social experiment where the assumption of no randomization bias cannot be tested from experimental data. Both approaches require assumptions that might be not verified without a specific collection of data, properly designed to test these assumptions.

When the control group is drawn from the population at large, even if strict comparability rules based on observable information are satisfied, which is really quite hard to achieve, it cannot be ensured that there are not differences in unobservables related to the program participation. This is the *econometric selection problem* as defined by Heckman (1979). In this case using the estimator (2.5) results in a non-identification problem since, abstracting from other regressors in the outcome equation, it approximates:

$$E[\hat{\alpha}_{ATE}] = \alpha + \{E[U_{it}|d_i = 1] - E[U_{it}|d_i = 0]\}.$$

In the case where $E[U_{it}d_i] \neq 0$ (*selection on unobservable*), $E[\hat{\alpha}_{ATE}]$ is expected to differ from α , unless the two final r.h.s. terms cancel out. Then, different estimators are needed.

The different methodologies for solving the evaluation problem with non-experimental data mainly depend on four factors: the type of information available to the researcher, the parameters of interest, the underlying model and how the counterfactual is constructed. As regards this, each class of estimators differs in the way it transforms

or adjusts the data, in order to estimate the counterfactual part. For the different approaches presented in this review, it will be considered how the various estimators construct this counterfactual and what kind of assumptions they require.

The following sections will present methods dealing with non-experimental data, characterizing the identification assumptions necessary to justify their application, possible reasons for their failure and empirical studies existing in literature. The focus is on the problem of identification, while the issue of precision or sampling variability is not discussed. Nevertheless, this is a fundamental issue for all practical applications; a reported effect estimate is absolutely worthless if it is not accompanied by a measure of variability.

Five distinct but related approaches existing in the program evaluation literature are considered. The next section starts with discussing the Instrumental Variable that is, together with the two-step Heckman selection estimator, closest to the structural econometric method, since it relies on the endogeneity problem. If longitudinal or repeated cross-section data are available, the diff-in-diff approach can be applied to obtain a more robust estimate of the treatment effects. The third approach, the so called matching estimator, requires detailed individual information for both participant and non-participant group, before and after the program. The intuition behind this non-parametric approach is to simulate with non-experimental data the randomize control of the experimental setting through independence assumption. When the participation rules is deterministic defined, the regression discontinuity design method provides a good way to deal with the selection bias between the two groups. In some sense the assignment rule here is the opposite to that in social experiment, but it turns out that it is good as random in a neighborhood of the critical value that defines the discontinuity point.

For each estimator I will discuss how the counterfactual part is estimated, together with the identification of the treatment impact in a homogeneous and heterogeneous setting, as well as the required assumptions, advantages and disadvantages.

2.5 Instrumental Variable Estimator

The main idea of this econometric approach is that one or more observable characteristics of individuals may well induce people to participate in a program, but have no direct consequences on outcome variables. In that way, a comparison of the average outcomes between participant and nonparticipant with different values on this characteristics can

replace the comparison of participants with a randomized control group. The correspondence between randomized experiment and instrumental variable approach can be found in Heckman (1996).

In a more formal way the IV method require the existence of at least one regressor, Z , which satisfies these two following conditions:

IV1 The participant status defined by the decision rule, conditional on X , is a non-trivial function of Z .

This means that $Pr(D = 1|X, Z)$ does not depend in a constant way on both X and Z , that is the instrument Z affects the decision rule independently from X ; in the specification of this rule the Z coefficients are non-zero. Thus, $E[D|X, Z] = P(D = 1|X, Z) \neq P(D = 1|X)$.

IV2 The regressor Z , conditional on X , is not correlated with the unobservable components (U^0, V) and (U^1, V) .

This implies that $E(U^0, Z) = E(U^1, Z) = E(V, Z) = 0$. This assumption means that the instrument Z has no influence on the outcome equation Y through the unobservable component, but it affects the outcome only trough the participation rule. In a homogeneous framework this means that only the level is affected by Z , while under heterogeneous setting the particular values of X determine how much is the influence of Z on Y .

These two assumptions together mean that the instruments Z provides a form of variation that is correlated with the decision rule but does not influence the potential outcome from treatment directly. This approach seems to be very easy to understand and also to implement. However in the treatment evaluation problem the instrument choice is not so easy to solve: it is not so trivial to find a variable that satisfies all the assumptions required to identify α . A possible solution is to consider lagged values of some determinant variable, when longitudinal data are available.

2.5.1 Homogeneous Treatment Effect

Under the above conditions, the treatment effect α is identified by applying the standard IV procedure; it is used only the part of the variation in D that is associated with Z . Thus, the instrumental variable estimator can be written as

$$\hat{\alpha}_{IV} = \frac{cov(y_i, Z_i)}{cov(d_i, Z_i)}$$

A special case arises when the instrument Z is a discrete variable that takes only the value 1 for a group of the observations and the value 0 for the remaining observations. In this case the instrumental estimator is equivalent to the following Wald estimator:

$$\hat{\alpha}_{IV}^{Wald} = \frac{\bar{Y}^1 - \bar{Y}^0}{\bar{X}^1 - \bar{X}^0}$$

where \bar{Y}^1 is the mean of Y across the observation with $Z = 1$ and \bar{Y}^0 is the mean of Y across the observation with $Z = 0$ and the analogous for X .

Alternatively, the effect of the homogeneous treatment might be found using both Z and X to predict D : a new variable \hat{D} is builded in order to be used in the regression instead of D . Another possibility is directly derived from the first assumption: it states that there must be at least two values of Z , for example Z' and Z'' , such that, for any X , $P(D = 1|X, Z') \neq P(D = 1|X, Z'')$. Moreover, applying the second assumption and equation (2.3), it might be used the law of iterated expectation to write:

$$E[Y|X, Z] = f^0(X) + \alpha P(D = 1|X, Z)$$

Then, the IV estimator is:

$$\hat{\alpha}_{IV} = \frac{E[Y|X, Z'] - E[Y|X, Z'']}{P(D = 1|Z') - P(D = 1|Z'')}$$

2.5.2 Heterogeneous Treatment Effect

Some problems might arise when the impact of a program is evaluated in a heterogeneous framework. To understand why, from equation (2.4) the error term is given by

$$U_{it}^0 + D_{it}(U_{it}^1 - U_{it}^0).$$

Even if the instrument Z is not correlated with U_{it} the same is not be true with respect to $U_{it}^0 + D_{it}(U_{it}^1 - U_{it}^0)$ because Z determines D_{it} by assumption. For this reason, conditions IV1 and IV2 are no longer enough to identify ATE or TTE. Then, some additional requirements on data must be imposed in order to identify a treatment effect in the heterogeneous framework. To be more clear, consider an example in which it is choose as instrument the distance between a unit's residence and the center in which is carried out the program. The assumptions stated above tell us that this distance

influences outcomes only through the participation indicator in the outcome equation. The problem in a heterogeneous setting arises because it is expected that people who live far away from the treatment location will participate in the program only when their expected gain from the program participation is relative large with respect to the lower expected gain of the people living closer, who incur less cost of participation. As a result, the post-program gains for all the participants also depend on how far away they live from the treatment location; thus the instrument is correlated with the unobservable component of the outcome. That is, knowing the distance between residences and program location tells us something about expected outcomes, which means that such distance is not a suitable instrument.

In the simplest case an additional assumption might be:

IV3 When deciding about participation, individuals do not use the information on the idiosyncratic component of the treatment effect $\alpha_i(X) - \alpha(X)$, where $\alpha(X) = E[\alpha_i(X)]$

If this assumption is satisfied, potential participants have no idea about future gains related to program participation, and their decision is taken on the basis of the average treatment effect. Then, together with the above assumptions IV1 and IV2, it might be identified the average treatment effect ($E[\alpha_i(X)]$) because

$$E[U_i^1 - U_i^0 | X, Z, D] = E[D_i[\alpha_i(X_i) - \alpha(X)] | X, Z] = 0.$$

On the other hand, selection on the unobservable is expected if units are aware of their idiosyncratic gains from treatment in a sense that individuals that gain more from participation are the most likely to participate within X -group. This selection process generates correlation between $\alpha_i(X)$ and Z .

Local Average Treatment Effect

A possible solution of the selection on unobservable problem described above is proposed by Imbens and Angrist (1994) who reinterpret the IV estimator as the effect of treatment from local changes of the instrument Z . They called it as the Local Average Treatment Effect (LATE). It depends on variation in an instrumental variable that is external to the outcome equation. However, opposite to the IV approach discusses so far, different instruments determine different parameters. To understand better what

LATE parameter wants to estimate, consider the example discussed before; if the distance to the nearest treatment location is the instrument, LATE estimates the effect of variation in this distance on the outcome variable of units who are induced to change their participant status as a consequence of the different costs, due to distance that is variable within a specified interval.

The basic idea behind this estimator is that some local changes in Z can reproduce random assignment, because agents take different decisions as they face different conditions uncorrelated to potential outcomes. However another assumption is needed to ensure that the two groups are comparable:

IV4 The participant status defined by the decision rule, conditional on X , is a non-trivial monotonic function of Z .

To define the LATE parameter more precisely, suppose D is an increasing function of Z . In a case where Z changes from $Z = z$ to $Z = z'$, where $z' > z$, units that modify their participant status, as a consequence of the variation in Z , are those that participate under $Z = z'$ excluding the ones that prefer to participate under $Z = z$ or, equivalently, those that do not participate under $Z = z$ excluding the ones that decide to not participate under $Z = z'$. Then, it can be estimated the expected outcome for treated and non-treated units for those influenced by the variation in Z as:

$$\begin{aligned} E[Y_i^1 | X_i, D_i(z) = 0, D_i(z') = 1] &= \\ &= \frac{E[Y_i^1 | X_i, D_i(z') = 1]P[D_i = 1 | X_i, z'] - E[Y_i^1 | X_i, D_i(z) = 1]P[D_i = 1 | X_i, z]}{P[D_i = 1 | X_i, z'] - P[D_i = 1 | X_i, z]} \end{aligned}$$

$$\begin{aligned} E[Y_i^0 | X_i, D_i(z) = 0, D_i(z') = 1] &= \\ &= \frac{E[Y_i^0 | X_i, D_i(z) = 0]P[D_i = 0 | X_i, z] - E[Y_i^0 | X_i, D_i(z') = 0]P[D_i = 0 | X_i, z']}{P[D_i = 1 | X_i, z'] - P[D_i = 1 | X_i, z]} \end{aligned}$$

Then, the estimated local average treatment effect is,

$$\begin{aligned} \alpha_{LATE} &= E[Y_i^1 - Y_i^0 | X_i, D_i(z) = 0, D_i(z') = 1] \\ &= \frac{E[Y_i | X_i, z'] - E[Y_i | X_i, z]}{P[D_i = 1 | X_i, z'] - P[D_i = 1 | X_i, z]} \end{aligned}$$

Since its dependence on the particular value of Z used, the LATE parameter does not represent TTE or ATE and it is different from the IV estimator discussed before. LATE

measures the impact of the treatment for units that move from $Z = z$ to $Z = z'$. In general, they do not represent the whole population, or the whole treated population. Thus, LATE represents a treatment effect on a sub-group of the treated units who are at the margin of participating for a given $Z = z$. In this sense, it can be viewed as the discrete approximation of the Marginal Treatment Effect (MTE), defined as the limit of the LATE estimator, when $z' - z \rightarrow 0$, that is for an infinitesimal variation in Z . The MTE represents the TTE for units that are indifferent between participating and not participating at $Z = z$ and can be written as,

$$\hat{\alpha}_{MTE}(X_i, z) = \frac{\partial E[Y|X_i, Z]}{\partial P[D = 1|X_i, Z]} \Big|_{Z=z}$$

All the three parameters (ATE, TTE and LATE) can be viewed as averages of MTE over different subset of the Z support: ATE is over the entire support of Z , TTE excludes the subset of the Z support when the treatment does not occur and LATE is over an interval of Z , where the size of the interval is define by two different participation rates.

2.5.3 Instrumental variable approach in practice

An example, among the several empirical studies that have been carried out to estimate a treatment effect with an IV approach in a program evaluation context, is the analysis of Angrist and Krueger (1991). The principal question the study wants to answer is on how compulsory school attendance affects schooling and earnings. Considering the education as the treatment and the length of education as an endogenous participation decision, the authors want to search for an instrumental variable that has to be correlated with an individual's participation decision and to have no independent effect on the outcome of interest, that is the weekly wage.

Considering the fact that most of the US school districts do not admit students to first grade unless they will attain the age of 6 by January 1st of the academic year in which they enter school, students born early in the year are older when they start school than students born later in the year. As a consequence, under any compulsory school leaving age, they can leave earlier and get less education. Interaction of school entry requirements and compulsory schooling laws compel students born in certain months to attend school longer than students born in other months. For that reason the authors proposed as instrumental variable the season of birth, which generates

exogenous variation in education that can be used to estimate different effects on some outcomes of interest, such as the impact of compulsory schooling on education and the effect of education on earnings.

The first step of the analysis is to recognize an evidence of the impact of the quarter birth on the level of education. The second step regards the estimation of the rate of return to education. Authors propose different models: the first one finds OLS and Wald estimates of the return to education, using the instrument variable as a dummy, equal to one for units born in the first quarter of the year. Another economic model is specified excluding the instrument from the wage equation and adding birth years dummy variables. The two models provide very similar estimates and the final interpretation of the results is that compulsory schooling laws are effective in compelling some students to attend school. Furthermore, those compelled to remain in school earn higher wages as a result of their extra schooling. Some final considerations are made about the problem that arise when instruments are only weakly correlated with the endogenous explanatory variables. This can lead to a large inconsistency in the IV estimates. That is why it is very important to examine the power of instruments in the first-stage regression. Bound *et al.* (1995) present evidence that weak instruments are a potential concern in this work of Angrist and Krueger (1991).

Another application of the IV approach can be found in Levitt (1997). The focus here is on the effects of police on city crime rates. Noting that increases in the size of the police force in large cities are disproportionately concentrated in election years, the proposed instrument is the timing of mayoral and gubernatorial elections. The identifying assumption is that timing of elections affects the size of the police force but does not directly affect crime rates.

To estimate how much profits of a firm's are captured by workers Van Reenen (1994) starts considering that, under imperfect competition in labor market, wages depend on company's profits as well as individual profitability (e.g. oligopoly and union bargaining). However, a regression of firm average wages on firms average profitability will be biased downwards because a wage shock will lead to lower profits. The solution he proposes is to use observed technological innovations as an instrument for profits. The underlying assumption is that this instrument should increase profits, but no direct affect wages.

2.6 Heckman Selection model

To understand how this method works let consider the model

$$\begin{aligned} Y_{it} &= X_{it}\beta_0 + \beta_1 D_i + U_{it} & \text{with } t > k \\ Y_{it} &= X_{it}\beta_0 + U_{it} & \text{with } t \leq k \end{aligned}$$

and the index model of the individual's non-random participation decision,

$$\begin{aligned} I_i &= Z_i\gamma + V_i \\ D_{it} &= 1 \quad \text{if } I_i > 0 \quad \text{and } t > k \\ D_{it} &= 0 \quad \text{otherwise} \end{aligned}$$

The main objective of this approach is to remove the sample selection problem that remains even after controlling for other determinants in the outcome equation; when $E[U, V] \neq 0$, estimating the outcome equation using OLS yields biased and inconsistent estimates of the treatment effect. The baseline idea of the Heckman Selection model, also called “*Heckit*” (Heckman (1979)), is to directly control for that part of the error term (U) in the outcome equation that is correlated with the participation dummy variable (D). Then, it is necessary to impose additional structure on the model:

- Z is exogenous in the outcome equation, $E(U|X, Z) = 0$
- X is a strict subset of Z . This implies there is at least one additional regressor in the participation decision
- the jointly density of the distribution of the errors U_{it} and V_i , $h(U_{it}, V_i)$, is known or it can be consistently estimated.

As a consequence of the additional assumptions, this method is more robust than the IV estimator. As above the homogeneous treatment effect is considered first.

2.6.1 Homogeneous Treatment Effect

To estimate the treatment effect the procedure uses two steps: first, it is estimated the part of the error term U_{it} which is correlated with the dummy variable D_i . Then it is included in the outcome equation and in the second step the effect of the program is estimated. By construction, what remains of the error term in the last equation of the outcome is not correlated with the participation decision.

The last assumption about the knowledge of the error terms joint distribution is necessary to obtain the estimation in the first step. To be more clear take the example in where U_{it} and V_i are assumed to follow a joint normal distribution and, for simplicity, consider the standardization $\sigma_V = 1$. Taking the expectations in the outcome equation and using the properties of the normal distribution, the conditional outcome expectation can be written as

$$\begin{aligned} E[Y_{it}|D = 1] &= X_{it}\beta_0 + \beta_1 + \rho \frac{\phi(Z_i\gamma)}{\Phi(Z_i\gamma)} \\ E[Y_{it}|D = 0] &= X_{it}\beta_0 - \rho_{(U,V)} \frac{\phi(Z_i\gamma)}{\Phi(Z_i\gamma)} \end{aligned}$$

where the final term on the right-hand of side of each equation corresponds to the expected value of the error term (U), conditional on the participation variable (D). This new regressor is the part of the error term which is correlated with the decision process. Including it in the outcome equation controls for non-random selection, enabling us to identify the treatment effect, separating the true impact of treatment from the selection process. The term $\phi(Z_i\gamma)/\Phi(Z_i\gamma)$ is the inverse Mills ratio, that is the quotient between the standard normal pdf and standard normal cdf evaluated at $Z_i\gamma$. To obtain an estimation of β_1 , the Heckman selection estimator for the selection model, the first step is to use observations for treated and non-treated units to estimate a probit model of participation, regressing I on Z . In the second step observations are used to estimate the outcome equation using OLS; in that way the estimated coefficients are consistent and approximately normally distributed.

One important advantage of this procedure in the homogeneous framework is that it is robust to choice-based sampling. This is the situation when the non-randomness is obtained drawing the comparison group of the non-treated from the population (see Heckman and Robb (1985)). Even if, as usual, the sample portion of the treated is not equal to the population one, and so, the treatment group is likely to be over-represented in the sample, robustness is achieved always by controlling the part of the error term U_{it} which is correlated with the participation variable D_i , since the remaining error term is orthogonal to D_i .

2.6.2 Heterogeneous Treatment Effect

When a heterogeneous framework is imposed, the outcome equation, abstracting from other regressor, takes the form

$$Y_{it} = \beta_0 + \beta_{1i}D_i + U_{it} \quad \text{when } t > k$$

when β_{1i} is the treatment impact of the i -th individual. Let $\bar{\beta}_1$ the population mean impact, ε_i the individual's deviation from the population mean and β_{1T} the mean impact of treatment on the treated. Thus

$$\begin{aligned} \beta_{1i} &= \bar{\beta}_1 + \varepsilon_i \\ \beta_{1T} &= \bar{\beta}_1 + E[\varepsilon_i | D_i = 1] \end{aligned}$$

where $E[\varepsilon_i | D_i = 1]$ stands for the mean deviation of the impact among participants. The outcome regression can now be rewritten as

$$Y_{it} = \beta_0 + \beta_{1T}D_i + \{U_{it} + D_i(\varepsilon_i - E[\varepsilon_i | D_i = 1])\} = \beta_0 + \beta_{1T}D_i + \xi_{it}$$

Now, the two steps procedure necessary to estimate the treatment effect requires knowledge of the joint density of U_{it} , V_i and ε_i . As before, take the example of a joint normal distribution and, for simplicity, consider the standardization $\sigma_V = 1$. Thus,

$$E[\xi_{it} | D_i = 1] = \text{Corr}(U_{it} + \varepsilon_i V_i) \text{Var}(U_{it} + \varepsilon_i)^{1/2} \frac{\phi(Z_i \gamma)}{\Phi(Z_i \gamma)} = \rho_{(U, V, \varepsilon)} \frac{\phi(Z_i \gamma)}{\Phi(Z_i \gamma)}$$

$$E[\xi_{it} | D_i = 0] = \text{Corr}(U_{it}, V_i) \text{Var}(U_{it})^{1/2} \frac{-\phi(Z_i \gamma)}{1 - \Phi(Z_i \gamma)} = \rho_{(U, V)} \frac{-\phi(Z_i \gamma)}{1 - \Phi(Z_i \gamma)}$$

Then, the outcome equation that provide a consistent estimation of the treatment effect becomes:

$$Y_{it} = \beta_0 + D_i \left[\beta_{1T} + \rho_{(U, V, \varepsilon)} \frac{\phi(Z_i \hat{\gamma})}{\Phi(Z_i \hat{\gamma})} \right] + (1 - D_i) \rho_{(U, V)} \frac{-\phi(Z_i \hat{\gamma})}{1 - \Phi(Z_i \hat{\gamma})} + \lambda_{it}$$

2.6.3 Selection models in practice

An example of an empirical study that applies the Heckman selection model for an estimation of a treatment effects, is the work of Willis and Rosen (1980); they want to measure the effect of the college education on earnings. Taking the college education as the treatment, the aim of this study is to find the relationship between individual characteristics and earnings, after controlling for selection bias. A second question they want to answer is if alternative earning prospects, as opposed to family background and financial constraints, influence the decision to attend college.

The starting point of the work is the specification of an econometric model of the college education decision constituted of two equations: one for the outcome (earnings) and the other one for the choosing level of schooling. After specifying a model for the expected values of earnings stream they define a college selection equation. The estimation procedure starts from this last equation; after that, the inverse Mills ratio are evaluated and included when structural earnings and growth equations are estimated. Finally the predicted values generated from the structural earnings and growth equations are included in the college selection equation to estimate the structural relationship between earnings and schooling.

2.7 Difference-in-Difference Estimators

A widely used approach to the evaluation problem when longitudinal or repeated cross-section data on nonparticipants in different periods are available, is the difference in difference approach, also called *diff-in-diff*. The main idea here is to use the additional time dimension to refine the counterfactual estimate and, thus, to reduce the selection bias in treatment effect estimation. This method is also called the *natural experiment* approach: the idea is to look for a natural occurring comparison group in order to mimic the properties of the control group in the properly designed experimental.

To understand how this method works, suppose information is available for a pre- and post-program periods, denoted respectively by t_0 and t_1 , ($t_0 < k < t_1$). In this case, the mean effect of the program on the treated units can be easily defined as:

$$\alpha = E[Y_{it1}^1 - Y_{it0}^0 | D_i = 1] - E[Y_{it1}^0 - Y_{it0}^0 | D_i = 1]$$

where the first expected value is the one of the pre-post intervention growth of the outcome registered in the treated units, while the second expected values refers to the

counterfactual before-after program growth always for the participants. The effect α cannot be directly estimated because the counterfactual growth is not observable. What can be calculated is instead the effect obtainable as:

$$\alpha^* = E[Y_{it1}^0 - Y_{it0}^0 | D_i = 0] - E[Y_{it1}^0 - Y_{it0}^0 | D_i = 1]$$

which yields unbiased estimates only if the expected value of pre-post program growth of Y , for the nonparticipants, corresponds to the counterfactual growth of treated units. This is the main assumption underlying the diff-in-diff estimator. To be more precise,

$$\bullet E(Y_{it1}^0 - Y_{it0}^0 | D_i = 1) = E(Y_{it1}^0 - Y_{it0}^0 | D_i = 0)$$

It means that participants and nonparticipants have the same mean change in the no-program outcome measures.

Another way to represent this approach is to express it in a regression term, rewriting the model (2.4) as follows,

$$Y_{it} = f_t^0(X_i) + \alpha_{it}(X_i)D_{it} + (\phi_i + \theta_t + \varepsilon_{it}) \quad (2.6)$$

where the error term U_{it}^0 , is being decomposed on an individual specific fixed effect, ϕ_i , a common macro economic effect, θ_t , and a temporary individual specific effect ε_{it} . The assumption above can be rewritten as,

$$E[U_{it}^0 | X_i, D_i] = E[\phi_i | X_i, D_i] + \theta_t$$

that is, selection into treatment is independent of the temporary individual specific effect ε_{it} . Then, it is easy to define the diff-in-diff estimator that measures the excess outcome growth for the treated compared to the non-treated. It can be written as,

$$\alpha_{DID}(X) = [\bar{Y}_{t1}^1(X) - \bar{Y}_{t0}^1(X)] - [\bar{Y}_{t1}^0(X) - \bar{Y}_{t0}^0(X)]$$

where \bar{Y} stands for the mean outcome among the specific group considered.

In the case of heterogeneous treatment is easy to demonstrate that the diff-in-diff estimator recovers the TTE since

$$E[\hat{\alpha}_{DID}(X)] = E[\alpha_i(X) | D_i = 1] = \alpha_{TTE}(X).$$

That is, the effect of treatment on the treated is identifiable, but not the population impact.

In the homogeneous effect case, one may omit the covariates from the equation of the diff-in-diff estimator and average over all the treated and non-treated units.

The principal drawbacks and weaknesses of the natural experiment approach arise from assumptions the method relies on. The first one is that there are no systematic composition changes within each group. This is related to the fact that there is no control for unobserved temporary individual specific components, ε_{it} , that influence the participation decision. The second assumption is that there are common time effects across groups. However, it is not so unlike that some macro effects have differential impact across the two groups. Some characteristics that distinguish between the treatment and the comparison group could not be equal and made the groups react different to common macro shocks. That is why an extended version of the diff-in-diff estimator has been studied, which allows to consider different trends between the two groups.

2.7.1 Trend Adjusted Diff-in-Diff

The basic principle upon which this approach bases on is that the availability of a series of observations $Y_{i,t-k-1}, Y_{i,t-k-2} \dots$ registered at times prior to the program intervention, allows one to control for pre-intervention differences between the treated units and the excluded ones. In particular, if it is available the information at a time $t - k - 1$, the size of the selection bias can be reduced; this third temporal observation allows one to be more precise on the estimate of the counterfactual growth and on the difference between the pre-intervention growth rate recorded for the treated units and the pre intervention growth rate registered for the excluded units. This difference is then used to correct the estimation of the counterfactual part that would be obtained with only two temporal observations.

To be more precise and to express the problem in a regression term, suppose that there are heterogeneous time effects between the treatment and control groups. Thus, the error term can be rewritten as,

$$U_{it} = \phi_i + \omega^D \theta_t + \varepsilon_{it}$$

when ω^D specifies the differential macro effect across the two groups. Then, the new assumption takes this form:

- $E[U_{it}^0 | D_i] = E[\phi_i | D_i] + \omega^D \theta_t$

Now, the diff-in-diff estimator identifies

$$E[\hat{\alpha}_{DID}(X)] = \alpha_{TTE}(X) + (\omega^1 - \omega^0)(\theta_{t1} - \theta_{t0})$$

It is easy to understand that this estimator represents the true TTE only when $\omega^0 - \omega^1 = 0$. To obtain a consistent estimate of the treatment impact, allowing for heterogeneous time effect, Bell *et al.* (1999) proposed the trend-adjusted diff-in-diff estimator, which takes the form,

$$\hat{\alpha}_{TADID} = [(\bar{Y}_{t1}^T - \bar{Y}_{t0}^T) - (\bar{Y}_{t1}^C - \bar{Y}_{t0}^C)] - [(\bar{Y}_{t''}^T - \bar{Y}_{t'}^T) - (\bar{Y}_{t''}^C - \bar{Y}_{t'}^C)]$$

where T and C refer to the treatment and control group and (t', t'') is another time interval, with $t' < t'' < k$ over which a similar macro trend has occurred. To be more precise, the authors specify that they require a period for which the macro trend matches the term $(\omega^1 - \omega^0)(\theta_{t1} - \theta_{t0})$. It is plausible that the most recent cycle is the most appropriate, with the minimum number of different effects across the target and comparison group.

Finally it is important to note that the assumption behind this approach is significantly weaker than the one needed for the diff-in-diff estimator without trend-adjustment; it is not required that the treated and the non-treated units have equal expected growth, over the pre-post program period, in the counterfactual case in which the treated individuals would not have been treated. The trend-adjusted estimator imposes that these two expectations have to be equal only after being corrected by the pre-intervention difference in the growth trends.

In a similar way, this procedure can be extended to a larger number of temporal observations improving the correction of the growth trends used to estimate the counterfactual part. As a consequence, the estimator that can be implemented require increasingly weaker conditions to yield unbiased estimates.

2.7.2 Difference in difference in practice

An application of the diff-in-diff estimator can be found in the work of Duflo (2001), which refers to a program of school construction launched in 1973 from the Indonesian government. To incentivate schooling and education and to increase the average number of schools, more than 61,000 primary schools were constructed.

As regards the treatment variable, exposure of an individual to the program was

determined by the number of schools built in his region of birth and his age when the program was launched. The question the work wants to answer is about the effect of the Indonesian school construction program on education and earnings. The diff-in-diff methodologies applied here refer to the differences between the region and the cohort of birth.

The starting point of the study is the specification of a model for the return to education on earnings. The main assumptions the model relies on regards some considerations about ages and regions of birth of the children. In particular, since Indonesian children normally attend primary school during age 7-12, all children born in 1962 or earlier were 12 or older in 1974, when the first schools were constructed, and hence were not affected by the program. On the other hand, for younger children, exposure to the program is an increasing function of their date of birth.

The diff-in-diff approach is applied to find the causal effect of the program, distinguishing between children with a little or no exposure to the program and children exposed to the entire time in primary school.

Another application of the diff-in-diff approach can be found in Eissa and Liebman (1995) who examine the impact of U.S. Earned Income Tax Credit for single women with children. The aim is to evaluate how the EITC reforms affected single women parents' labor supply. The authors suggest to use as control group the population of single women without children, because this group was not eligible for the EITC. The final results are that labor force participation rises for the treatment group after the reform, but there are no significant response in annual hours and annual weeks from reform.

2.8 Matching Estimators

The matching estimator tries to solve the problem of identifying the treatment effect on outcomes, in a non parametrical way. Like the diff-in-diff method, it is a general approach; that is, it does not require any particular specification of the outcome equation or of the participation decision. Furthermore, it does not require the additive specification of the error term and any exclusion restrictions. The fact that it is not a parametric method makes it quite flexible in the sense that it can be combined with other methods to obtain more precise estimates. On the other hand, it assumes that analysts have access to a set of conditioning variables, X ; that is, generous and good quality meaningful data is needed.

The main idea of this method is to replicate the condition of an experiment in a case where the data are non-experimental and a no randomized group is available. This is possible by constructing a correct sample counterpart for the no available information on the outcomes had they not been treated. Each participant is paired with one or more members of the non-treated group on the basis of the X values. Then, the main purpose is to match treatment and comparison units that are similar in terms of their observable characteristics. Then, it is possible to directly compare treated and non-treated units to obtain an estimation of the impact of the intervention, because the only remaining difference between the two groups is the program participation, like in the total random assignment case. Thus, the assumption necessary for the matching estimators regards the existence of a set of variables such that, conditioning on them, the counterfactual outcome distribution of the participants is the same as the observed outcome distribution of the non-participant. To be more formal the matching method is based on the following assumption:

M1:

$$Y^0, Y^1 \perp D | X.$$

It means that the non-treated and treated outcomes are independent of the participation status, conditioning on the set of variables X . As a consequence, the distribution of outcomes $F(Y^0|D = 1, X) = F(Y^0|D = 0, X) = F(Y^0|X)$ and $F(Y^1|D = 1, X) = F(Y^1|D = 0, X) = F(Y^1|X)$. In that sense, the available regressors “adjust” the differences between treated and non-treated units.

This assumption was first presented in this form in Rosenbaum and Rubin (1983) who called it as “ignorable treatment assignment” or “unconfoundedness”. Lechner (1999) refers to this as the “conditional independence assumption”, while in the work by Barnow *et al.* (1980) it is referred to as “selection on observables” in a regression setting; this is why, given X , the non-treated outcomes are what the treated outcome would have been had they not been treated, that is selection occurs only on observable.

To see the link with standard exogeneity assumptions, consider the case of constant treatment effect: $\alpha = (Y_{it}^1 - Y_{it}^0)$ for all i . Furthermore, suppose that the outcome of the control group is linear in X_i :

$$Y_{it}^0 = \beta_0 + X_i' \beta_1 + \varepsilon_{it} \quad \text{with} \quad \varepsilon_{it} \perp X_i$$

Then we can write

$$Y_{it} = \beta_0 + \alpha D_{it} + X_i' \beta_1 + \varepsilon_{it}$$

Given the constant treatment effect, unconfoundedness assumption is equivalent to independence of D_{it} and ε_{it} conditional on X_i . That is, D_{it} is exogenous. Note that, without the constant treatment assumption, unconfoundedness does not imply a linear relation with independent errors, in mean.

Given the assumption above, it is possible to find for each treated observation Y^1 a non treated (set of) observation(s) Y^0 with the same X realization. That is why the method is called matching: it tries to match each observation with another, or more, which is similar. If the assumption holds, the outcome of the control group is the required counterfactual. In that sense this approach tries to re-building an experimental data set.

To ensure that this assumption has empirical content, it is also necessary to assume that there are participants and non-participants for each X for which is possible to make a comparison. Then, it is necessary to impose another assumption to guarantee the existence of the required counterfactual. To be more precise,

M2:

$$0 < P(D = 1|X) < 1.$$

It means that all treated units have a counterpart on the population of the non-treated and anyone is a possible participant. In a finite sample, of any size, we replace this condition by the empirical probability. It means that, this last assumption does not ensure that it holds in any sample. It is a strong assumption, especially where programs are direct to a specific groups. Rosenbaum and Rubin (1983) called it as “strong ignorability”. It has important practical consequences for program evaluation in the sense that failure to satisfy it appears to be one of the most important reason why matching methods produce biased estimates of the impact of a program. For some comments on the plausibility of these assumption see Imbens (2003).

Under the two assumptions above, it is possible to create a comparison group that replaces an experimental control group in one key respect: conditional on X the distribution of the counterfactual outcome Y^0 for the participants is the same as the observed distribution of Y^0 for the comparison group.

In the literature there has been some controversy about the plausibility of these two assumptions in economic settings. The main debate regards the possibility of dependence between the potential outcome and the choices taken by the agents in

order to optimize their behavior, even if conditional on covariates. To a more detailed description of these remarks see Imbens (2003).

However, it could be important to note that the assumptions of the matching approach do not imply no selection bias; indeed, as long as the means exist, they imply that

$$E[Y^0|X, D = 1] = E[Y^0|X, D = 0]$$

and that

$$E[Y^1|X, D = 1] = E[Y^1|X, D = 0]$$

This does not imply $E[U^0|X, D = 1] = 0$, i.e. no selection bias. Instead, matching balances the bias, like experiments:

$$E[U^0|X, D = 1] = E[U^0|X, D = 0] = E[U^0|X].$$

To compute the method of matching outcomes in the treatment group, denoted with Y^T , are matched with the outcome of a sub-sample of persons in the comparison group, Y^C , to estimate a treatment effect. Given the common support, that is the set of all possible values the vector of explanatory variables X may assume, individual gains from the program among the subset of participants who are sampled and for whom one can find a comparable non-participant, must be integrated over the distribution of observables among treated units and re-scaled by the measure of the common support, called S^* . Thus, the matching estimator for the ATT is the empirical counterpart of

$$\frac{\int_{S^*} E[Y^T - Y^C|X, D = 1] dF(X|D = 1)}{\int_{S^*} dF(X|D = 1)}.$$

This result represents the expected value of the program impact: it is the simple mean difference in outcomes over the common support S^* , weighted by the distribution of participants.

It is worth to note that to identify the TTE, the independence assumption might refer only to the non-treated units. Thus, it can be written $Y^0 \perp D | X$. On the other hand, the assumption $Y^1, Y^0 \perp D | X$ is necessary for the identification of the ATE parameter.

In practice, to construct matches, it is needed a measure of distance between the units with respect to the X , in order to define the units in the comparison sample who are neighbors to each treated units i . Heckman *et al.* (1997b) present different kind

of alternative matching schemes proposed in literature. Here, only the most common methods are introduced.

One simple algorithm to identify the most similar comparison units to be matched to the treated units is the *nearest neighbor matching*, developed by Rubin (1973). In this procedure for each unit i is selected only one “most similar” unit j , chosen among the group of the non-participant units. To select the comparison unit some distance metric must be minimized,

$$\min_{j \in N_c} \|X_i - X_j\|$$

where N_c is the size of the subsample of the comparison units and $\|\cdot\|$ is a metric measuring the distance in the X characteristics space. The most widely used one is the Mahalonobis distance, where the metric used to define the neighborhoods for i is

$$\|\cdot\| = (X_i - X_j)' \Sigma_c^{-1} (X_i - X_j)$$

where Σ_c^{-1} is the covariance matrix in the comparison sample. As a result, the two groups, the comparison one and the group of the treated units, have the same size. Depending on the common support between the groups, two different versions of the nearest neighbor matching can be considered: the *nearest available matching* and the *matching with replacement*. The main difference is that the latter allows to many treated units to be matched with the same excluded units. In this way, each participant will be matched even when just few excluded units are comparable to the treated individuals, because of the small common support.

Another possible procedure is the *radius matching*, which admits for each treated unit to be matched with more than one excluded unit (see Dehejia and Wahba (1998a), also for a comparison between the different algorithms). In this procedure the matches are made only if

$$\|X_i - X_j\| < \delta$$

where δ is a tolerance level chosen by the evaluator. Otherwise unit i is bypassed and no match is made to this individual. Similarly to the matching with replacement this method allows a given excluded unit to be matched more than one time.

If one wants to use the entire comparison sample, a possible solution is the *kernel matching*, which is a smooth method that reuses and weights the comparison group sample observation differently for each treated unit i with a different X_i . Let $W(i, j)$ the weight placed on observation j in forming a comparison with observation i , with

this algorithm the weights are

$$W(i, j) = \frac{K(X_j - X_i)}{\sum_{j=1}^{N_c} K(X_j - X_i)}$$

where K is a kernel function.

In general, independently of the algorithm used, the form of the matching estimator is determined by computing the mean difference across the i . Thus,

$$\hat{\alpha}_M = \sum_{i=1}^{N_T} \left\{ Y_i - \sum_{j=1}^{N_C} W(i, j) Y_j \right\} w_i \quad (2.7)$$

where $W(i, j)$ is the weight for individual i with respect to the comparison observations j , and w_i takes into account the re-weighting that reconstructs the outcome distribution for the treated sample. To be more clear, where the nearest neighbor algorithm is used, the estimator is given by

$$\hat{\alpha}_M = \sum_{i=1}^{N_T} (Y_i - Y_j) \frac{1}{N_T}$$

where j is the nearest neighbor non-treated unit to i . More efficient estimators use also the variance to construct the weight of the observations (see Heckman *et al.* (1997b) and Heckman *et al.* (1998b)).

The main idea of the matching approach is to ensure the suitable set of observable characteristics X is being used to obtain the correct counterfactual. As in the specification of conventional econometric model, there is the same uncertainty about which X to use. Heckman *et al.* (1997b) discuss some tests in order to choose the appropriate X regressors. Furthermore, in practice, it is really hard to find a similar control unit if a very detailed information is available, because of the more restricted common support. It is evident that there is a trade-off between the share of the common space and the information to use. If however the correct amount of data is used, the only reason the treatment effect, conditional on X , might not be identified, is selection on unobservables. It is important to note, however, that some problems can arise where there is a situation of non-overlapping support of X and an incorrect weighting over the common support. To be more precise, the bias term, due to the difference between treated and

non-treated, can be decomposed in three different components:

$$E[Y^C|X, D = 1] - E[Y^C|X, D = 0] = B^1 + B^2 + B^3, \quad (2.8)$$

where B^1 is the bias due to the non-overlapping support of X and B^2 represents the error part related to the misweighting on the common support of X . These two sources of bias might be corrected through the matching process of choosing and reweighting observations. The third term B^3 is the econometric selection bias resulting from selection on unobservables, that is assumed to be zero.

Another way to use the matching methods it is the so called *regression-adjusted matching* proposed by Rubin (1979) and deepened in Heckman *et al.* (1997b) and Heckman *et al.* (1998b): the main idea is to compute a regression of the outcome on the X regressor and to use the regression-adjusted Y_i , computed as $R(Y_i) = Y_i - X_i\beta$ in place of Y_i in the above calculations.

A further development of the matching method is to use it in a parametric approach; in general it does not require functional form assumptions for the outcome equation, but, if a functional form is maintained it is possible to implement the matching method using regression analysis (see Barnow *et al.* (1980)). To obtain an estimation of the effect of the treatment, it is estimated first the relationship between the outcome and the observable X for the treatment and control groups. After that, the predicted outcomes are used to do a comparison between the two groups in order to obtain an estimation of the impact. One advantage of this method is that it does not require the common support condition for the distribution of X , that might be very different in the treated and comparison group. The comparability is achieved imposing the functional form.

2.8.1 The Propensity score

The use of the matching methods, like all the non-parametric methods, is seriously limited if the dimensionality of the vector X is high. In practice, it could be difficult to find some control units with similar values of X , if X is composed by a large vector of variables. A solution can be to match on a function of the X . The most common and useful is to compute the so called propensity score, $p(x)$, which is probability of participation,

$$p(x) = P(D_i = 1|X_i).$$

It can be interpreted as the probability that a unit i is selected for treatment, given its values of X . Thus, it summarizes in a single parameter the impact of all the observable pre-intervention units' characteristics that differentiate the treated units from the excluded ones.

The propensity score were proposed by Rosenbaum and Rubin (1983) who demonstrate that the conditional independence assumption remains valid controlling for $p(x)$ instead of X :

$$Y^0, Y^1 \perp D \mid p(x).$$

Conditioning on $p(x)$ reduces the dimension of matching problem down to the matching on the scalar $p(x)$.

As in the standard matching, when using the propensity score, the comparison group for each treated units is chosen with a pre-defined measure of distance; once defined the neighborhood for each units, the second choice regards the appropriate weights to associate each selected individual of the control group with the treated one. The several proposed solutions are the same presented above: from weight equal to one to the nearest observation and zero for the others, to equal weights to all, otherwise, kernel weights.

One of the most important consideration that has to be done on the propensity score is about its estimation; an application of the matching method with the use of the propensity score needs a demonstration that a suitable model for $p(x)$ has been selected. Rosenbaum and Rubin assumes that $p(x)$ is known rather than estimated. For some comparison between the two different situations see Heckman *et al.* (1998b): they present the asymptotic distribution theory for the kernel matching estimator both in the case where the propensity score is known and the case in which it is estimated, parametrically and nonparametrically. On the other hand a study of Hahn (1998) shows that $p(x)$ is ancillary for the estimation of ATE, but its knowledge may improve the efficiency of the TTE estimation by reducing the "dimensional" problem.

The propensity score can also be used as a control variable in an outcome regression, i.e. conditioning on a propensity score method or also to stratify the data sample based on similar pre-intervention characteristics, that is data stratification on a propensity score. In the first case, the predicted probability $\hat{p}(x)$ is added to the outcome equation; as argued in Rosenbaum and Rubin (1983) this is a convenient way to deal also with non-linearities in the relationship between outcome variables and pre-intervention characteristics of units. The data stratification on $p(x)$ method can be adopted to evaluate program, separating data into strata based on the units' propensity score and

to estimate the mean differences between treated and non-treated individuals of each stratum (see details in Dehejia and Wahba (1998a)).

Finally in the *weighting* approach the propensity score is used as a weight to create balanced samples of treated and control observations; units are weighted by the inverse of the probability of receiving treatment and an estimation of the effect of the program is simply

$$\frac{1}{N} \sum_{i=1}^N \left(\frac{D_i Y_i}{p(x)} - \frac{(1 - D_i) Y_i}{1 - p(x)} \right).$$

For more details together with some combinations of these methods see Imbens (2003).

2.8.2 Matching Diff-in-Diff Approach

The assumption behind the matching approach is a really strong one if individuals can decide according to their forecast outcomes. To overcome this drawback the matching is combined with the diff-in-diff approach, in order to control for an unobservable determinant of participation (see Blundell *et al.* (2001)). However, it is worth to note that what follows is valid as long as this unobservable component can be represented by separable individual and/or time-specific components of the error term. To be more precise let consider the model (2.6) specified above in a matching framework. The independence assumption, conditional on the set X , takes the form

$$(\varepsilon_{t_1} - \varepsilon_{t_0}) \perp D \mid X$$

where t_0 and t_1 stands for the before-after program period, ($t_0 < k < t_1$). The idea behind this method is that only the individual-specific changes require additional control, since the diff-in-diff controls for the other determinants. This assumption implies that control units have changed their outcomes from a pre- to a post-program time in the same way treatment units would have done had they not be treated. Furthermore, this is true both for the observable component and the unobservable time trend. The estimator of the treatment effect (2.7) can now be extended and rewritten as

$$\hat{\alpha}_{MDID} = \sum_{i \in T} \left\{ [Y_{it_1} - Y_{it_0}] - \sum_{j \in C} W_{ij} [Y_{jt_1} - Y_{jt_0}] \right\} w_i.$$

It is obvious that longitudinal data are required. However, in the case where this type of data are not available, the matching diff-in-diff can be extended for the repeated cross-section data case. One needs to implement matching three times for each treated units after being treated: once to find the comparable treated units before the program, and the second and third time to find the control before and after the intervention. If the same assumptions hold, the estimation of the treatment effect can now be computed by

$$\hat{\alpha}'_{MDID} = \sum_{i \in T_1} \left\{ \left[Y_{it_1} - \sum_{j \in T_0} W_{ijt_0}^T Y_{jt_0} \right] - \left[\sum_{j \in C_1} W_{ijt_1}^C [Y_{jt_1} - \sum_{j \in C_1} W_{ijt_0}^C Y_{jt_0}] \right] \right\} w_i$$

where T_0, T_1, C_0 and C_1 stand for the treatment and comparison groups pre- and post-intervention and W_{ijt} is the weight attributed to unit j in the respective group, treated or control, and time t when comparing with the treated individual i .

2.8.3 Matching approach in practice

One of the most important empirical study dealing with the application of the matching method is the work of Blundell *et al.* (2001). The aim of the study is to estimate the effect of a mandatory job assistance program in the UK, the “New Deal for the Young Unemployed” that helps young unemployment people make their way into or back to work. The program is addressed to all young people aged 18- 24, whose were claiming Job seekers allowance for 6 months and thus were unemployed at least in the previous 6 months. For 4 months they are intensively monitored and given job search assistance. If they still have not any job then they can get employer wage subsidy or, alternatively a period, of training or full-time education.

The program started in January 1998 only for a tree-month experimental period, when it was carried out in 12 regions, until April, when the program was launched in the whole UK. The scope was to perform an experiment with the first 12 regions to obtain a counterfactual for the rest of the UK.

The analysis of the authors deals with the impact of this program on employment in the first 18 months of the scheme. In particular, it measures the effect on the probability of moving into a job during the 4 month assistance period of job search, conditional on 6 months as unemployment. Since the program was targeted to a specific age group, a natural comparison group would be similar individuals with the same unemployment status but slightly too old to be eligible. Using the diff-in-diff estimator,

a before and after comparison can be made; to improve the estimates also a matching diff-in-diff approach was implemented in the study. Thanks to the pilot area, the diff-in-diff approach has two possible comparison groups: the areas, because New Deal was launched in some pilota areas for 3 months, and the ages. It means that one can compare 18-24 year old individuals with 6 months unemployment in pilot areas versus 18-24 year old people with 6 months unemployment in non-pilot areas. Otherwise the comparison can be carried out with reference to the age: the 18-24 group in pilot areas is comparable with the 25-30 year old group in pilot areas.

Another important study that reveals the power of the matching approach is the analysis of Heckman *et al.* (1997b). It evaluates matching under different assumption on the richness of available information. Data collected from the Job Training Partnership Act (JTPA) was used to examine the empirical performances of matching methods by comparing the parameter estimates from randomization with those from non-experimental matching methods. They consider a variety of non-experimental control groups, such as eligible non-participants individuals (resident in the same narrow geographical region) and no shows, that were experimental persons assigned to treatment who enrolled in JTPA but dropped out before receiving services.

A more recent study by Smith and Todd (2005a) is based on the same data collected from the JTPA, that are also used from LaLonde (1986) to asses the reliability of the non-experimental methods by comparing the results with the ones obtained using experimental data. The results reveal that matching may improve the results when only cross-section data are available. Obviously, the choice of the variable to use for the match plays a fundamental role. However, where longitudinal data are available, the quality and the precision of the estimates improve independently of the method chosen. The discussion on these themes is not yet ended: the study of LaLonde (1986) has encouraging a famous discussion on the effectiveness of the matching estimators between Smith-Todd and Dehejia (see Dehejia (2005) and the reply of Smith and Todd (2005b)).

2.9 Regression Discontinuity Estimators

Regression discontinuity design (RDD) constitutes a special case of “selection on observables”. In fact, the essential element of this model, originally introduced by Campbell and Stanley (1963), is that the probability of assignment to treatment depends in a discontinuous way on some observable variable S . That is, participants are assigned to

program solely on the basis of an established cutoff score on a pre-intervention measure. To better understand, consider the case in which a set of units willing to participate are divided into two groups, according to whether the pre-program measure is above or below a specified threshold. Those who score below the threshold are excluded to the intervention, while those who score above are exposed. Note that, in some sense, the assignment rule to treatment is the opposite here to that in random assignment; it is a deterministic function of some observable variables. But, it will turn out that, assignment to treatment is as good as random in the neighborhood of the discontinuity.

The reason that distinguishes it from randomized experiments, and from other quasi-experimental strategies, is its unique method of assignment. This cutoff criterion implies the major advantage of RDD: first, it is appropriate when we wish to target a program to those who most need or deserve it. Second, it is certainly more attractive than a non-experimental design in the sense that in a neighborhood of the threshold the RDD presents some features of a pure experiment. Moreover, other features distinguish this method from the standard selection on observable and reveal all the powerful of this approach. First, there is a common support for participant and non-participants. Thus RDD is an attractive procedure when there is selection on observables but the overlapping support condition required for matching breaks down. Second, the selection rule is deterministic and known by assumption.

On the other hand, the design has two main limitations: the first one regards the fact that its feasibility is only confined to those cases in which selection takes place on an observable pre-program measure. Secondly, even when it is feasible, it only identifies the mean effect at the discontinuity point for selection. If we are in a case with heterogeneous treatment effects, it tells us nothing about units away from the threshold. In this sense, RDD is able to identify only a local mean impact (LATE).

To understand clearly how the method works, let consider before the similarities between a randomized experiment and a RDD. As stated before, the most attractive property of the randomized experiment is that the impact of the program is simply the difference between the mean outcomes for treated and non-treated units. Although the RDD lacks random assignment of individuals, it shares some important features with this experimental approach. If S is the variable, or the set of variables, according to which units are selected into the program, and \bar{s} is the threshold for selection, the dummy variable D takes value 1 only if units' score is above \bar{s} . Then,

$$D = \mathbb{I}(S \geq \bar{s})$$

where \mathbb{I} is the indicator function. Thus, the probability to be treated, conditional on S , steps from 0 to 1 as S crosses the threshold \bar{s} . This represents the so called *sharp* RDD, introduced by Trochim (1984): the treatment D is known to depend in a deterministic way on some pre-program observable continuous variable S , $D = f(S)$, and the point \bar{s} , where the function $f(S)$ is discontinuous, is assumed to be known.

An alternative case, that is more general of the sharp design, is the so called *fuzzy* RDD, where D is a random variable given S , but the conditional probability of receiving treatment, $Pr(D = 1|S)$ is known to be discontinuous at \bar{s} . This design differs from the previous case in that the treatment assignment is not a deterministic function of S ; there are some variables which are unobserved by the evaluator that determine the assignment rule. For details see Hahn *et al.* (2001). One example of the fuzzy RDD arises when units do not comply with the mandated status, dropping out of the program or seeking alternative treatments.

The common feature of the two designs is that the probability of receiving treatment, $Pr(D = 1|S)$, viewed as a function of S , is discontinuous at \bar{s} . It constitutes the first assumption of the method. To be precise:

RDD1:

i) the limits

$$\begin{aligned} D^+ &\equiv \lim_{S \rightarrow \bar{s}^+} E[D|S = s] \text{ and} \\ D^- &\equiv \lim_{S \rightarrow \bar{s}^-} E[D|S = s] \text{ exists} \end{aligned}$$

ii) $D^+ \neq D^-$.

In both cases the main idea is that conditioning on S allows to identify the average effect of the program in a neighborhood of the cutoff point \bar{s} , that is a local version of the mean impact of the intervention.

The mean treatment effect at the point \bar{s} is identified if it is satisfied the following assumption:

RDD2:

$$E[Y^0|\bar{s}^+] = E[Y^0|\bar{s}^-]$$

Then, the mean value of Y^0 conditional on S is a continuous function of S at \bar{s} . This condition for identification requires that in the counterfactual world, no discontinuity

takes place at the threshold for selection. This condition allows to identify only the average impact for subjects in a right-neighborhood of \bar{s} , that is the effect of treatment on the treated (ATT). The identification of the effect of treatment on the non-treated, requires a similar continuity condition on the conditional mean $E[Y^1|S]$. In practice, it is difficult to think of cases where condition RDD2 is satisfied and the same condition does not hold for Y^1 .

From the second assumption it follows:

$$(Y^0, Y^1) \perp D|S = \bar{s}.$$

Because of this property the RDD is referred to as a quasi-experimental design.

An estimate of the impact of the program can be obtained under different assumptions regarding the heterogeneity of the effects among units. If a common treatment effect among individuals is supposed, together with the assumption that in the absence of treatment persons close to the cutoff point are similar, the effect of the treatment can be written as

$$\alpha^{RDD} = \frac{Y^+ - Y^-}{D^+ - D^-}$$

where $Y^+ \equiv \lim_{S \rightarrow \bar{s}^+} E[Y|S = s]$ and $Y^- \equiv \lim_{S \rightarrow \bar{s}^-} E[Y|S = s]$.

In the sharp design, $D^+ = 1$ and $D^- = 0$; hence the common treatment effect is identified by

$$\alpha_s^{RDD} = Y^+ - Y^-$$

Then, in the simple case of a constant treatment effect the regression line at the cutoff point represents the effect of the program. For example, under the hypothesis of linearity in the relationship between the set of variables S and the output, α can be estimated without bias by OLS estimation of:

$$Y = \beta_0 + \alpha D + \beta_1 S + U$$

On the other hand, if it is considered the case of variable treatment effects among units, other assumptions have to be imposed, to generalize the identifying strategy followed above:

- $E[\alpha_i|S = s]$ regarded as a function of S , is continuous at \bar{s}
- D is independent of α_i conditional on S near to \bar{s} .

Then an expression for the mean treatment effect might be:

$$E[\alpha_i|S = \bar{s}] = \frac{Y^+ - Y^-}{D^+ - D^-}.$$

As before, with a sharp design it is identified by

$$E[\alpha_i|S = \bar{s}] = Y^+ - Y^-$$

It is worth to note that, the second assumption regarding the conditional independence implies that units do not select into treatment on the basis of expected gains from exposure. This assumption, as viewed before, is often invoked in the literature of program evaluation, but it may be considered unrealistic in a situation where individuals self-select into treatment. Furthermore, in this case of heterogeneous treatment, identification was possible by comparison units close to the threshold \bar{s} who did and did not receive treatment. This means that the effect can only be identified at $S = \bar{s}$.

For both designs, to obtain a consistent estimation of the parameter of interest it is sufficient replace the ratio with consistent estimators of $\hat{y}^+, \hat{y}^-, \hat{x}^+, \hat{x}^-$. They may be computed with different methods, adopting parametric or non-parametric approach. For details see Hahn *et al.* (2001).

2.9.1 Regression discontinuity Design in practice

An example of an empirical study that applies the regression discontinuity design for an estimation of treatment effects, is the work of van der Klaauw (2002): it evaluates the effect of financial aid offers of colleges and university on student enrolment decision. The work shows how discontinuities in an East Coast college's aid assignment rule can be exploited to obtain credible estimates of the aid effect without having to rely on arbitrary exclusion restrictions and functional form assumptions.

Following this work, another important study analyzes the effects of tuition fees on graduation time adopting a RDD approach: it is the working paper proposed by Garibaldi *et al.* (2007). They base their empirical analysis on detailed administrative data from Bocconi University in Milan, a private institution that, during the period for which they have information (1992-2000), offered a 4-years college degree in economics. This dataset is informative on the question under study not only because more than 80% of Bocconi graduates typically complete their degree in more than 4 years, but also because it offers a unique quasi-experimental setting to analyze the effect of the

tuition profile on the probability of completing a degree beyond the normal time. Upon enrollment in each academic year, Bocconi students in the sample are assigned to one of 12 tuition levels on the basis of their income. A RDD is used to compare students who, in terms of family income, are immediately above or below each discontinuity threshold. These two groups of students pay different tuitions to enroll, but should otherwise be identical in terms of observable and unobservable characteristics determining the outcome of interest, which is the decision to complete the program in time.

Chapter 3

The continuous treatment case

3.1 Introduction

The objective of this chapter is to give an overview of the literature on program evaluation in a continuous treatment setting. It is not so rare to find in practice situations where treatment regimes need not to be binary and units might be exposed to different levels or doses of treatment. This could be true both in economic or medical applications. In these situations studying the impact of such treatment as if it were binary can mask some important features on it.

Our intention is to build a basic statistical framework for our research, starting by an evaluation of the current studies on this topic. After a brief introduction on the continuous treatment setting, the focus will be on the analysis of the relevant literature provided by different authors on this issue. The most important studies on program evaluation with a continuous treatment are presented and analyzed in order to find the common characteristics, the main advantages and drawbacks of all approaches that will constitute the starting point of our analysis.

3.2 From binary to continuous treatment

Most of the relevant literature on the program evaluation deals with the estimation of causal effects of a binary treatment on one or more outcomes of interest in a non-experimental framework. In practice, however, treatment regimes need not to be binary and individuals might be exposed to different levels or doses of treatment. Then, in these

situations, it can be meaningful to use the information of the treatment level to estimate different kinds of treatment effects as a function of the doses. In other words, studying the impact of such treatment as if it were binary can mask some important features on it. Moreover other parameters might be of interest. When a binary treatment is evaluated the main focus is on the estimation of an average treatment effect; in a continuous setting many parameters might be important and meaningful. For example, it could be interesting learning about the form of the entire function of average treatment effects over all possible values of the treatment levels. In other words, one might be interested in studying how the effects change when the level of the treatment changes. More, another interesting parameter might be the “optimal” dose: optimal in the sense that it is the treatment level that maximizes the average effects. In other cases one could be interested in the derivative of the average effects or in knowing if there is a level at which the curve of the effects has a discontinuity point or a “turning” point.

In the last ten years the interest on the generalization of the program evaluation framework from a binary treatment setting to a more general structure for the treatments has increased rapidly. The most important reason is perhaps the fact that, in more and more implementations of public policies or interventions, there are cases with a more complicate structure than an easy situation where there are only treated and non-treated units. The most relevant cases might be classified in two groups: the multiple and the continuous treatment programs. The first group includes all the cases in which the policy consists of a variety of different programs. An example might be active labor market policies, which comprehend job-search assistance, training programs, public employment interventions, wage subsidies etc. Another case, that belongs to the multiple treatment setting, is when there are different discrete levels of treatment. An example is the evaluation of the effects of the years of schooling on individual earnings. The specified model might distinguish the impact of many different education levels, thus allowing the attainment of different educational qualifications to have separate effects on earnings. On the other hand, there are many applications of public policies that includes a strictly continuous treatment. It is the case, for example, of a firm incentives programs, that consist of a series of subsidies given to firms in order to achieve some employment or business growth goals. Or again, the case of a medical treatment to some patients which consists of different doses of a medicinal.

In general, the non-binary models would seem a more attractive frameworks since a wide range of treatment levels with potentially very different effects might be of interest. However, even if cases of multiple or continuous treatment represent a generalization of the binary treatment framework, they have some particular features that make them

very different. Here, the focus is on the continuous treatment setting: the idea is to study the relation between the effects of a policy and the treatment levels, identifying and estimating the possible parameters of interest.

The common approach followed from the evaluation literature when analyzing a binary treatment is the potential outcome approach developed by Rubin (1974). It might be easily extended to the continuous case: just consider a random sample of units, indexed by $i = 1 \dots N$, and for each unit i the existence of a set of potential outcomes, $y_i(T)$, for $T \in [0, t_1] = \mathcal{T}$. To be more clear, let consider an observed value of the treatment level t : the potential outcome $y_i(t)$ represents the outcome unit i would received if exposed to treatment level $T = t$, where T takes values in the interval $[0, t_1]$. Each individual receives exactly one of these treatment levels; before participation in the policy, each potential outcomes is latent and could be observed if the individual received the respective program. Ex-post, i.e. after the policy, only the outcome corresponding to the dose the individual received is observed, that is $y_i(t_i)$. This extension of the potential outcome approach constitutes the basis of all the studies on the program evaluation in a more general treatment regime framework.

The following sections will present the most important works that deal with this topics. For each contribution, it will be briefly described the main ideas and the empirical applications, followed by some personal discussions on the main advantages, limits and possible developments. This analysis of the literature will constitute the starting point for the further chapter in which the real contribution on the topic of the program evaluation with continuous treatment will be presented.

3.3 An overview on the literature

Despite the interest on generalization of the program evaluation framework from a binary treatment to a more general setting for the treatments has been recognized it is not yet a topic that has been deeply defined and studied. Although it is rapidly increasing, it represents a particular branch of the program evaluation that should be better analyzed and developed, because it could reveal some important results.

As stated above, the common approach of the few works that deal with this issue follows the generalization of the potential outcome setting. Furthermore, there is another common characteristic that guides these analysis: it refers to the generalization of the propensity score approach of the binary treatment case. In fact, in all the studies,

even if in different ways, the objective is to develop the propensity score in a continuous setting in order to remove any bias associated with differences in observable and unobservable characteristic among units. Then, the few works presented in literature can be classified on the basis of the use of this propensity score or function. After the definition of this quantity, some studies use it in a parametric structure, while other follow a non parametric approach, such as a matching estimator. This distinction will guide and characterize this overview of the literature.

Another important classification might be done on the basis of the parameter of interest. Some works focus on the potential outcome at each level of treatment, $Y(T)$, while others have as objective the estimation of some treatment effects, as the difference of some potential outcomes at different treatment levels. This distinction will not guide our presentation. However, the parameters of interest will have an important role in the next chapter where a new methodological approach will be proposed. For that reason, the estimated quantities of each work will be well specified, in order to underline differences and potential developments of each alternative analysis.

3.4 Generalized propensity score: parametric approach

One of the first study that deals with the continuous treatment is the work of Imbens (1999): an extension of the propensity score methodology is proposed that allows for estimation of average casual effects with multi-valued treatments. This work represents the starting point for the next analysis (Hirano and Imbens (2004)) where the propensity score method is extended in a setting with continuous treatment. The key assumption is the generalization of the unconfoundedness hypothesis for binary treatment to the multivalued case:

$$Y(t) \perp T \mid X \quad \forall t \in [t_0, t_1]. \quad (3.1)$$

Next they define the *Generalized Propensity Score* (GPS) as $R = r(T, X)$, where

$$r(t, x) = f_{T|X}(t|x)$$

is the conditional density of the treatment given the covariates. Together with the balancing property of the GPS, i.e. $X \perp 1\{T = t\} \mid r(t, X)$, similar to that of the standard propensity score, the unconfoundedness assumption implies that assignment to

treatment is unconfounded given the generalized propensity score. Thus for every t ,

$$f_T(t|r(t, X), Y(t)) = f_T(t|r(t, X)).$$

They use this result in order to remove any biases associated with differences in the covariates. The estimation of the parameter of interest $\mu^t = E[Y(t)]$ is obtained with a two steps procedure:

- (i) $\beta(t, r) = E[Y(t)|r, (t, X) = r] = E[Y|T = t, R = r].$
- (ii) $\mu^t = E[Y(t)] = E[\beta(t, r(t, X))].$

For a practical implementation of the proposed methodology the authors discuss estimation and inference in a parametric version of this procedure. In this sense this work might be classified as one following a parametric approach: the estimation and inference problems are handled with a parametric function while the basic framework is more general, and nothing prevent to implement it with more flexible approaches, as stated by the authors.

In the first stage what they propose is to use a normal distribution for the treatment given the covariates:

$$r(T, X) = T_i | X_i \sim N(\beta_0 + \beta_1' X_i, \sigma^2)$$

The estimated GPS is:

$$\hat{R}_i = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \exp\left(-\frac{1}{2\hat{\sigma}^2}(T_i - \hat{\beta}_0 - \beta_1' \hat{X}_i)^2\right)$$

where β_0, β_1 and σ^2 are estimated by maximum likelihood.

In the second stage they model the conditional expectation of Y given T and R using a quadratic approximation:

$$E[Y_i | T_i, R_i] = \alpha_0 + \alpha_1 T_i + \alpha_2 T_i^2 + \alpha_3 R_i + \alpha_4 R_i^3 + \alpha_5 T_i R_i$$

These parameters are estimated by OLS using the estimated GPS \hat{R}_i . Finally the estimate average potential outcomes at treatment level t are obtained from:

$$\widehat{E[Y(t)]} = \frac{1}{N} \sum_{i=1}^N (\hat{\alpha}_0 + \hat{\alpha}_1 t + \hat{\alpha}_2 t^2 + \hat{\alpha}_3 \hat{r}(t, X_i) + \hat{\alpha}_4 \hat{r}(t, X_i)^2 + \hat{\alpha}_5 t \hat{r}(t, X_i))$$

To obtain an estimation of the entire dose-response function this expected mean can be estimated for each level of treatment one is interested in.

The last part of the paper refers to an application of the proposed method. The data set consists of individuals winning the Megabucks lottery in Massachusetts in the mid-1980's. The interest is on the effect of the amount of prize on subsequent labor earnings. The estimated average effect of the prize are obtained by adjusting the differences in background characteristics using the propensity score. To see whether this specification is adequate, the authors investigate how it affects the balance of the covariates. They discretize both the level of the treatment and the GPS. First they divide the range of the price into intervals and compute the quintiles of the GPS evaluated at the median of the price in each group. Then, for each covariates it is investigated the balance by testing whether the mean of the observations of each price interval with the GPS, evaluated at the median price, that belongs to each quintile is different from the mean of the observations in the same interval of the GPS but in the other treatment groups combined.

This work represents one of the most important theoretical contribution to the program evaluation with continuous treatment: it explicitly considers the effects of different levels on the outcome, trying to deal with the selection on observable issue by removing the bias associated with differences in the covariates. However, there are some considerations that might be noted. First of all, let consider the parameter of interest: in this work the focus is on the curve of the potential outcome at each level of treatment. In the traditional literature on the program evaluation the parameter of interest regards the estimation of an effect of the treatment. The theoretical construct of the potential outcomes is traditionally used in order to estimate an effect by comparing treated and non-treated units. Instead, in this work, a proper estimation of this kind of effects is not possible to obtain. Following the potential outcome approach of Rubin (1974), you need the information of the non-treated units in order to obtain some value of their potential outcomes. Instead, in this work any comparison between participants and non-participants is done: this is easy to understand if one considers that only the data of the treated units are used. Thus, what is possible to obtain following the approach of Hirano and Imbens (2004) is an estimation of the effects between units at different treatment levels, by comparing the values of the estimated potential outcome equation for different levels, but not an estimation of the effects between treated and non-treated units.

A direct consequence of this first consideration reveals another important matter: the selection rule is not considered. The paper does not make any distinctions between

positive treatment level and the treatment at the level zero, that is the non-treatment status. Non-treated units are ignored, and no considerations are made on which are the factors that might influence the treatment status versus the opposite situation of non-treatment.

Finally, the last personal consideration concerns with the practical implementation of the method where a parametric approach is applied. This might cause the classical problem of mis-specification of a model, related to the parametric assumption. Thus, as suggested by the authors, the proposed method might be applied following a non parametric approach.

3.5 Some non-parametric approaches

In this section the focus is on the studies following a non parametric approach: distinct paragraphs will regard the use of matching and subclassification estimators.

A non parametric approach for the evaluation of a public policy with a continuous treatment is proposed by the working paper of Flores (2004). The main focus here is on more parameters of interest than the common average treatment effects. In particular, the author focuses on three objects:

- $\mu^t = E\{Y(t)\}$ for all $t \in \tau$
the entire curve of average potential outcomes, or dose-response function.
- $\alpha_0 = \arg \max E\{Y(t)\}$
the treatment dose at which the curve is maximized
- $\mu(\alpha_0) = E\{Y(\alpha_0)\}$
the maximum value achieved by the curve.

The author points out the advantages of a non-parametric approach: it overcomes the problem of an arbitrary choice of a discretization of the treatment variable, as suggested by Royer (2003), and the sensitivity of the results to model specifications.

In the first part of the paper the three estimators are presented under the assumption of randomization of the treatment doses. After the author moves to the case of non-experimental data. In this setting the key assumption is again the unconfoundedness assumption. But here a stronger version, with respect to equation (3.1) proposed by Hirano and Imbens (2004), is used :

$$\{Y(t)\}_{t \in \tau} \perp T | X$$

The stronger form of independence is maintained in this work because in practice, as argued by the author, it can be difficult to find applications in which the weak assumption may be plausible but the stronger form may not. This hypotheses allows to control for systematic differences in the observed covariates across treatment doses because the dose-response function can be written:

$$\begin{aligned} E[Y(t)] &= E_X[E[Y(t)|X = x]] \\ &= E_X[E[Y(t)|T = t, X = x]] \\ &= E_X[E[Y|T = t, X = x]] \end{aligned}$$

This suggests a regression approach to estimate $E[Y(t)]$. The author proposes non-parametric estimators based on a kernel function (Nadaya-Watson multiple regression estimator, or nonparametric mean regression estimator). However, for calculation of the estimators under the independence assumption it is needed first to estimate the non parametric regression of the observed outcome on the treatment dose and the covariates. This may be a problem if the dimension of the covariates is large. To deal with this problem of “dimensionality”, the paper discusses the use of the GPS, such as it is presented by Hirano and Imbens. However, the author proposes a non parametric estimation of the propensity score using nonparametric kernel estimators. The estimators for the three parameters of interest are two-step nonparametric estimators, where the GPS is estimated nonparametrically in the first step. To be more specific, the three estimators are:

$$\hat{E}[Y(t)] = \frac{1}{n} \sum_{i=1}^n \lambda(\hat{r}_i) \hat{g}_h(t, \hat{r}_i) \quad \text{for all } t \in \tau$$

$$\hat{\alpha} = \arg \max_{t \in \tau} \frac{1}{n} \sum_{i=1}^n \lambda(\hat{r}_i) \hat{g}_{h_1}(t, \hat{r}_i)$$

$$\hat{E}[Y(\alpha_0)] = \frac{1}{n} \sum_{i=1}^n \lambda(\hat{r}_i) \hat{g}_{h_2}(t, \hat{r}_i)$$

where $\hat{g}(t, \hat{r})$ is the Nadayara Watson multiple regression estimator,

$$\hat{g}(t, \hat{r}) = \frac{\sum_{j=1}^n Y_j K(\frac{t-t_j}{h}, \frac{\hat{r}-\hat{r}_j}{h})}{\sum_{j=1}^n K(\frac{t-t_j}{h}, \frac{\hat{r}-\hat{r}_j}{h})},$$

$\hat{r}(t, x)$ is the nonparametric estimator of the GPS and $\lambda(\cdot)$ is a trimming function used

to avoid the “denominator problem” by keeping a denominator bounded away from zero.

Then, the author moves from the regression context and discusses other ways to use the GPS. For example in a matching framework. This approach will be discussed in the part of this section dedicated to the matching estimators. Another proposed solution is to adopt a weighting approach to estimate the dose-response function as a natural extension to the continuous treatment case of the weighting-by-the-propensity-score approach used in the literature when the treatment is binary.

Finally, the author illustrates the techniques developed in the paper by presenting an empirical application to estimate non-parametrically the turning point of the inverted U-type relationship between some indicators of environmental degradation and income per capita, known in this literature as the “Environmental Kuznets Curve”.

The main advantage of this work is the non parametric structure, which allows to relax any parametric assumption and to avoid any mis-specification problem. Moreover, the new parameters of interest proposed by the author are really meaningful and interesting.

On the other hand, as in the work of Imbens and Hirano, a proper estimation of an effect of the treatment is not obtained: the parameter of interest is always the curve of the potential outcome and, another time, no comparison between participants and non-participants is done. Also the consideration about the selection rule is the same as before: there are no distinctions between positive treatment level and the treatment at the level zero.

3.5.1 Subclassification on the propensity score

Another important contribution to the program evaluation with a continuous treatment that deals with the generalization of the propensity score approach is the work of Imai and van Dyk (2004). The aim of the analysis is to develop theory and methods that encompass all the techniques applied to the binary, ordinal or categorical treatment case and widen their applicability by allowing for arbitrary treatment regimes, categorical, ordinal, continuous, semi-continuous, or even multi-factored. This is done by developing the theoretical properties of the propensity function, which is a generalization of the propensity score of Rosenbaum and Rubin.

To evaluate the effect of the treatment, the authors rely on the two standard assumptions: the SUTVA and the standard conditional independence assumption between treatment and outcomes, given the observed covariates. The parameter of interest concerns the average potential outcome $Y(t)$ at each level t , that is the average over the

population of the covariates of the distribution $f(Y(t)|X)$. To overcome the problem of a mis-specification adopting a parametric approach to model the variable of interest the authors propose to use non-parametric techniques: matching and subclassification are commonly used. However, as the dimensionality of X increases, these methodologies become impossible in practice. Thus, they propose a generalization of the propensity score method. After defining the propensity function as the conditional probability of the actual treatment given the observed covariates, $f_\psi(T|X)$ where ψ parameterizes this distribution, this set of parameters ψ must be estimated because in practice is unknown. This parametric model defines the propensity function, $e_\psi(\cdot|X) = f_\psi(\cdot|X)$.

To simplify the representation of the propensity function, the authors make an assumption: there exists a finite dimensional parameter θ that uniquely represents $e_\psi(\cdot|X) = e[\cdot|\theta_\psi(X)]$. This implies that the propensity function depends on X only through $\theta_\psi(X)$, i.e. θ is sufficient for T . The main advantage of this approach is that the parameter θ is typically of much lower dimension than is X .

The methodological contribution of the paper follows after these definitions: the authors show that given the propensity function the conditional distribution of the actual treatment does not depend on observed covariates, that is the balancing score property, and that the strong ignorability assumption holds with X replaced by the propensity function. Then, it can be averaged $f(Y(t)|e(\cdot|X))$ over the distribution of the propensity function to obtain $f(Y(t))$ as a function of t . To accomplish this average two solutions are suggested: subclassification or matching. In this latter case, the authors refer on the work of Lu *et al.* (2001), that will be discussed in the following paragraph. They argue that, although matching methods may be useful in particular setting, they believe subclassification is a more generally applicable strategy because it allows for simpler implementation of more complex analysis models.

The subclassification solution implies that, once estimated $\hat{\theta}$ and compute for each observation $\hat{\theta} = \theta_{\hat{\psi}}(X)$, the observations with the same or similar values of $\hat{\theta}$ are subclassified into a moderate number of subclasses. Within each subclass $f(Y(t)|T = t)$ is modelled and the relevant causal effect is obtained, e.g. the regression coefficient of $Y(t)$ on t . Then, the average causal effect can be computed as a weighted average of the within-subclass effects. This last parameter describes how the full distribution of the potential outcome can be approximated at a particular level of the treatment. Although this full distribution is sometimes appropriate in practice more often it is summarized by its mean, $E[Y(t)]$: this is the approach the authors take in the example they present.

After presenting two Monte Carlo experiments to illustrate how controlling for the propensity function can improve the statistical properties of estimated causal effects,

the work ends with an application: the propensity function method presented is applied to two dataset; it is estimated the effect of smoking on medical expenditure and the effect of schooling on wages.

This work represents another important contribution to the literature on the propensity score approach with continuous treatment. What might be pointed out is that the continuous dimension is in some sense underestimated. The proposed generalized propensity function might be used to estimate the full distribution of the effects, but what the authors suggest is to use its mean to summarize it; in this way the information given by the continuous treatment is lost.

Another important consideration deals with the relation between the levels and the effects: the authors propose to regress the outcome variable on the treatment dose, within each subclass. This implies a proportional structure of the levels, that might be a strong assumption. Another specification of this relation might be adopted.

Moreover, as in the previous work, the selection rule is not considered. There are no distinction among the levels, in particular between non treated units and individuals treated at any level. This indirectly justifies the fact that only the data of treated units are used.

3.5.2 Matching estimators

This section will discuss that part of the non-parametric methods adopting a matching approach to solve the evaluation problem in a continuous treatment setting. This part of the literature plays an important role in our analysis where the empirical application will be presented.

A contribution on the continuous treatment is the empirical work of Behrman *et al.* (2004). The continuous dimension is given by the length of exposure of a treatment. The authors do the analysis in the context of studying the effect of a preschool development program targeted toward disadvantage children between the ages of 6 and 72 months in Bolivia on some child outcome measures related to health, psycho-social skill and cognitive developments. They mainly focus on estimation of two parameters of interest: the average treatment effect on the treated, as in the binary case, and what they call marginal program impacts, that are the effects of increasing the duration in the program.

In the first part of the paper they develop a model of enrollment that gives an economic interpretation for the average treatment effects that they estimate later. In order to control for potential bias due to non random selectivity into the program, they propose to use matching estimators allowing for a continuous dose of treatment.

After identifying the main assumptions of this approach, that they specify as in the binary case, keeping separated the group of the non treated from the set of the treated units, without make any consideration about the treatment level, they present the “cumulative” and “marginal” matching estimators. In the first case the parameter of interest is the average impact of treatment on the treated given by:

$$E(\Delta^{T0}|t > 0) = E[Y(t) - Y(0)|t > 0]$$

where $t \in \tau$ denote time spent in the program, with $t = 0$ for nonparticipants. The key identifying assumption they use is:

$$E[Y(0)|t_i = t, X = x] = E[Y(0)|t_i = 0, X = x], \quad \text{for all } t \in \tau.$$

A stronger version of this assumption is $Y(0) \perp t \mid X$ for all $t \in \tau$. The next step is the estimation of the expected values for participants and nonparticipants, conditional on X and on the level t of treatment, using local nonparametric regression methods. To be more specific they estimate:

$$\hat{E}[Y(0)|x_i, t_i = 0] = \sum_{k \in \{t_i=0\}} Y_k(0) W_k(||X_k - X_i||)$$

$$\hat{E}[Y(t_i)|x_i, t_i > 0] = \sum_{k \in \{t_i>0\}} Y_k(t_k) W_k(||t_k - t_i||, ||X_k - X_i||)$$

where $W_k(||X_k - X_i||)$ and $W_k(||t_k - t_i||, ||X_k - X_i||)$ are weights that add to one and come from the local nonparametric regression of $Y_k(0)$ on X , and of $Y_k(t_k)$ on t and X respectively, and $|| \cdot ||$ is the euclidean distance. In this way, as suggested by the authors, the weights depend on the distance between t_k and t_i , allowing the impact to depend on the duration of time in the program. They also point out that an alternative approach would be construct the weighted averages for the estimation of the conditional expectation for participants over the set of observations that are selected into the program and that receive a treatment level equal to t_i . Instead they do local averaging over durations t because there may not be many observation at any individual duration value. Thus, they emphasize one of the most relevant problem in the estimation of effects with a continuous treatment: due to the continuity of the treatment variable it would be difficult in practice to find observations with a level value

exactly equal to t .

Then, for the estimation of the average treatment effect on the treated they propose:

$$\hat{E}(\Delta^{T0}|t > 0) = \frac{1}{n} \sum_{i \in \{t_i > 0\} \cap \{t_i \in S_p\}} \{\hat{E}[Y(t_i)|x_i, t_i > 0] - \hat{E}[Y(0)|x_i, t_i = 0]\}$$

where S_p is the region of common overlapping support and n is the cardinality of the set $\{t_i > 0\} \cap \{t_i \in S_p\}$.

For the estimation of the second set of parameters, what they called the “marginal” estimators, the focus is on the marginal treatment effect of increasing duration in the program from a level to another, to t_0 to t_1 :

$$E(\Delta^{t_0, t_1}) = E[Y(t_1) - Y(t_0)|t > 0]$$

The way the authors propose to compute it is to use only the data on participants, drawing comparisons between program participants who have taken part in the program for different lengths of time. An advantage of this approach is that it does not require assumptions on the process governing selection into the program. On the other side, there is another potential source of non random selection: the process governing selection into alternative program duration. Again, matching methods can be used to solve this selection problem relating to the choice of program duration, under the assumption that units who have taken part in the program for different lengths of time can be made comparable by conditioning on observed unit with similar characteristics. The solution proposed is similar to the one presented before and the conditional expectations are estimated by the same local regression method as described above.

The final consideration, before presenting some empirical results, is on the possibility of a dimensional reduction: they assume that the conditional mean independence assumption holds with X replaced by the propensity score $p(x) = P(T > 0|X = x)$, where $T = \{t : t > 0\}$. Then, the conditional expectation can be estimated by three- and two-dimensional non parametric regression using the distance across the propensity scores instead of the values of the covariates.

A final theoretical consideration regards the selection on unobservable issue: the necessary assumptions for the matching estimators proposed are not likely to be satisfied if unobservables that are related to outcomes are important determinants of program selection. One option is to use a difference-in-difference matching strategy that allows for time-invariant unobservable differences in the outcome between participants and

non-participants. However the data used in the analysis do not allow application of this estimator, because program participants are only observed after they already entered the program. Only the marginal impact of short versus long durations might be estimated using this estimator, allowing selection to be based on unobservables.

This work is one of the most important application of matching estimators in a continuous treatment setting: it explicitly considers the continuous dimension of the treatment and tries to develop and to modify the existing matching estimator in order to estimate causal effects and some relation between effects and treatment length. Moreover, it gives some importance to the selection process. In the first part of the paper it is presented a theoretical model for the enrollment decision, and it is recognized a selection process composed of two elements: the program participation and the alternative program durations. However, the proposed theoretical model for the enrollment decision, is not the real selection process, but rather a useful interpretation for the treatment impact estimates. As before, there are no distinctions across the levels and between non treated and treated units. On the other hand, this is not relevant in the marginal effects estimation case, as suggested by the authors. Instead, if the parameter of interest is the estimation of the treatment effects when the counterfactual is no treatment the assignment rule has a central role that might not be excluded from the analysis.

As regards the dimensional reduction, it is important to note, as pointed out by Flores (2004) that the independence assumption does not follow from the conditional mean independence stated before. Or, it is not the case that $Y(0) \perp t | X$ implies $Y(0) \perp t | p(X)$ for all $t \in T$. The assumption about $p(X)$ made by the authors has no relation to the unconfoundedness-given- X assumption discussed in the works of Hirano and Imbens (2004) and Flores (2004).

Another important consideration has to be done with respect to the final estimates presented in this work. In the first part of the analysis the focus is on the importance of the continuous dimension of the treatment. However, what the authors propose for the estimation of causal effects is an average treatment impact, that is a weighted average impact of participating in the program relative to not participating for the treated units. In this way all the information relative to the continuity of the treatment is lost and no relation between treatment level and doses is obtained.

Another relevant work that deal with the estimation of causal effects using a matching approach is the study of Lu *et al.* (2001). The focus is on the evaluation of the effects of a media campaign launched in the United States intended to reduce illegal

drug. Since the campaign was implemented throughout the country there is no unexposed or control group available for use in evaluating the effect of the program. Hence, in this case, only the data of the participants are available and, as a consequence, only marginal effects might be estimated. The main idea of the authors is to compare units who receive different exposures to the campaign, but who were similar in terms of baseline characteristics. For that reason, they propose a matching approach.

Multivariate matching with doses of treatment differs from the usual treatment-control matching mainly in two ways. First, pairs must not only balance covariates, but also must differ markedly in dose, in such a way that the final high- and low-dose group have similar or balanced distribution of observed covariates. Second, any two subjects may be paired, so that the matching is nonbipartite, that is within a single group. In this case the group is given by the treated units. Finally, a propensity score with doses must be used in place of the conventional propensity score. Then, this different approach affects three aspects of matching: the definition of the propensity score, the definition of distance and the choice of optimization algorithm.

After discussing the relationship between the authors' approach and that of Imbens (1999), it is presented the optimal matching algorithm used to minimize total distance and the definition of propensity score used in order to take in account the continuity dimension of the treatment variable. The key issue is to use a model that allows the distribution of doses given covariates to depend on these regressors via a scalar function of the covariates. This happens in McCullagh's ordinal logit model (McCullagh (1980)), that is used in the work to obtain a balancing score, used in the matching. Finally, it is presented the particular distance used: the goal of matching with doses is not only to balance the observed covariates, but also to produce pairs with very different doses. The authors propose a distance measure that decreases both as the propensity scores become similar and as the assigned treatment become dissimilar.

This work is another example of an extension of the matching approach to a continuous treatment case: as before, only the data of treated units are used and the selection process is not considered. In this case, however, there should be no other choices, because of the overall cover of the program. For that reason, the estimated parameter of interest is not the impact of the campaign against the benchmark of no program, but rather the marginal treatment effect of increasing exposure in the program from a low to a high dose. In some sense, the continuous dimension of the treatment is reduced to a comparison between two groups, such as in the traditional program evaluation case with a binary treatment, with the difference that the comparison is across units that receive some treatment. Again no relations between the doses and the treatment effects

are estimated.

The matching approach adopted by the authors of this work is different from the other studies briefly summarized in this review. The relevant distinction is on the role of the continuous treatment variable in the matching procedure: here the authors stress the fact that matching has to be done between units with similar covariates but very different treatment dose, such as in the binary treatment case when the matches are computed between the group of the treated and non treated individuals. On the other hand, the work of Behrman *et al.* (2004) and the one of Flores (2004), point out that the matching has to be done between similar units, also in the treatment level received. They justify this assumption arguing that matching is informative about the potential outcomes, or the effects, if it is done by comparing units that receive a dose sufficiently close to the level one is interested in. However, it is important to note that the works are different also in the estimated parameters of interest and this might be justify the different approach adopted.

As mentioned before, also the work of Flores (2004), presented above, discusses the use of the generalized propensity score with a matching approach. He points out that, since the continuity of the treatment it would be difficult in practice to find observations with a dose value exactly t . Thus the matching has to be done not only on the GPS, but also in the treatment level. A reasonable way in which this method could be done is by matching observation on $||t - T_j, r(t, X_i) - r(T, X_j)||$, with $|| \cdot ||$ being a given metric. A disadvantage is that one can end up predicting the dose-response function at t using observation that received doses which are very far from t . As a consequence, the author proposes another method to match the units. Consider a window of size δ_n around t , where as usual δ_n is a sequence of positive real numbers tending to zero as $n \rightarrow \infty$. Then observed outcomes with $T_i \in [t - \delta_n, t + \delta_n]$ can be thought as an approximation to the potential outcomes of those observations at t . For observations with $T_i \notin [t - \delta_n, t + \delta_n]$ one looks for matches based on the GPS to impute their missing potential outcomes at the t level. Then the matching estimators can be written as

$$\hat{E}[Y(t)] = \frac{1}{n} \sum_{i=1}^n \hat{Y}_i(t)$$

with

$$\hat{Y}_i(t) = \begin{cases} Y_i, & \text{if } T_i \in [t - \delta_n, t + \delta_n] \\ \frac{1}{M} \sum_{l \in S_M(i)} Y_l, & \text{if } T_i \notin [t - \delta_n, t + \delta_n] \end{cases}$$

where $S_M(i)$ is the set of indices for the M closest matches for unit i in terms of

$|r(t, X_i) - r(t, X_j)|$, with $i \neq j$ and $T_j \in [t - \tilde{\delta}_n, t + \tilde{\delta}_n]$ ($\tilde{\delta}_n$ is a sequence tending zero as $n \rightarrow \infty$). Note that in this way the matching is done by comparing units that receive a dose sufficiently close to t in order for them to be informative about the potential outcomes at t .

3.6 Conclusions: our starting point

The aim of this review was to understand the state of the literature on program evaluation with continuous treatment. From this, some important personal considerations and comments might be done, in order to better understand the motivations which will guide the next work. In particular they deal with:

- comparison between treated and non-treated units;
- the participation decision process;
- the parameters of interest.

The first issue is the most relevant also because the other two depends on it. To be more precise, it regards the type of comparison one is interested in. The studies presented above mainly focus on the comparison among treated individuals. On the contrary, what we are interested in regards a comparison between units treated at different levels and non-treated units. It follows that the selection process that identifies participants versus non-participants, becomes a fundamental source of non-random selection in the identification of any policy effects, together with the process governing participation into alternative program doses.

As regards the quantity to estimate, there are two important consideration that might be pointed out. First, when the focus is on estimation of the policy effects, there is often an underestimation of the continuity dimension; the information given by the continuous treatment variable is lost because what is estimated is an average effect among different levels of treatment. On the other side, when the estimation is a function of the doses, the parameter of interest is the potential outcome $Y(t)$ rather than some policy effect.

Starting from these considerations, the main question our analysis want to answer is: why do not focus on the policy effects on the treated versus the non-treated units at different treatment level? Thus, the idea is to compare participants with non-participants, in order to estimate the effects of an intervention for each level of treatment one is interested in. Furthermore, why do not consider how these effects are related with the

continuous treatment variable? The idea is to find some relation among treatment levels and treatment effects.

The next chapter will deal with these issues. A new methodological approach will be proposed: it refers to a development of matching estimators. The choice of this kind of estimators, rather than some alternative approaches, comes from their well known good properties in the binary treatment case (see Heckman *et al.* (1998b) and Heckman *et al.* (1997b)).

Chapter 4

A new approach to empirical estimation

4.1 Introduction

As mentioned before, the previous chapter regarding a review on the main contributions on the continuous treatment program evaluation constitutes the starting point for this part of the work. Following the main limits, drawbacks and considerations discussed before, the aim is to present a different estimate approach on the topic of the program evaluation with continuous treatment.

The idea is to study the relation between the effects of a policy and the treatment levels, identifying and estimating the possible parameters of interest and trying to solve the main inconsistencies of the analysis discussed before. Starting from a specification of the continuous treatment setting, it will be presented a new specification for the selection process, the parameters of interest and the new matching approach adopted to a new empirical estimation of the treatment effects. The main differences and developments with respect to the previous literature will be underlined in the course of the analysis.

4.2 The continuous treatment setting

As briefly mentioned before, the common approach followed from the evaluation literature is the potential outcome approach: $y_i(T)$ represents the set of potential outcomes,

for each unit i , given a random sample indexed by $i = 1 \dots N$, and T represents the continuous variable indicating the treatment level. As in the binary case, the definition of potential outcome already made implicit uses the assumption of the stable-unit-treatment-value (SUTVA), that is no interference between units. It assumes that all the potential outcome $y_i(T)$ of individual i are not affected by the allocation of other individuals to the treatments. Thus, it is assumed that the observed outcome y_i depends only on the treatment level to which individual i is assigned and not on the allocation of other individuals. For each unit i there is also a vector of covariates X_i and the level of treatment received, $t_i \in [0, t_1] = \mathcal{T}$. Thus, the observed information are the vector X_i , the treatment received t_i and the corresponding potential outcome, $y_i = y_i(t_i)$.

Then, with a continuous treatment variable, each unit is characterized by a set of potential outcome. Following what is traditionally done in the literature on binary treatment, the sets of potential outcome may be divided in two groups: $y_i(T)$ for all the outcomes under the active treatment level T , with $T \in]0, t_1]$, and $y_i(0)$ otherwise. The difference with the traditional binary case is that, in the continuous setting, the set of active treatments $y_i(T)$ includes an infinite number of potential outcomes, depending on the treatment variable T .

The outcome Y is assumed to depend on a set of observable covariates X and on the participation status. The equations of the potential outcomes for individual i can be generically represented as:

$$\begin{aligned} y_i(t_i) &= f^T(X_i, t_i) + u_i(t_i), \quad \text{for } T > 0 \\ y_i(0) &= f^0(X_i, t_i) + u_i(0), \quad \text{for } T = 0 \end{aligned}$$

The functions $f^T(\cdot)$ and $f^0(\cdot)$ represent the relationship between the set of covariates X and the potential outcome $Y(0)$ e $Y(T)$, while the terms $U(0)$ and $U(T)$ identify the mean zero error terms, assumed to be uncorrelated with the X . This vector of variables is assumed known at the participation decision time, and not influenced by treatment.

The missing data problem, that is the impossibility of observing units under all the treatment status, is here more complicated, because the treatment status are not more only two, but are infinite. The general observed outcome Y can be written as:

$$y_i = d_i y_i(t_i) + (1 - d_i) y_i(0)$$

where D is a dummy variable indicating the treatment status: in particular $D = 1$ if the individual has been treated and $D = 0$ otherwise, and $y_i(t_i)$ is the particular

potential outcome at the observed level t_i . In practice, when $D = 1$ we observe $y_i(t_i)$, when $D = 0$ we observe $y_i(0)$.

4.2.1 The selection process

Selection into treatment determines both the treatment status d_i and the treatment level t_i ; then it may be considered as composed of two processes. The participation decision assignment will determine the treatment status d_i , while the treatment level process will determine the dose t_i . For simplicity, the participation assignment will be stated as the first processes. It is worth noting, however, that they occur at the same time. Moreover, it is assumed they occur together at a fixed moment in time and depends on the information available at that time. This information is summarized by a set of observable variables $\mathbf{X} = \{\mathbf{W}, \mathbf{Z}\}$ and unobservable $\varepsilon = \{V, U\}$, where \mathbf{W} identifies the treatment status and \mathbf{Z} the subsequent treatment level. Assignment to treatment is then assumed to be made on the basis of

$$t_i = \begin{cases} g(Z_i) + u_i & \text{if } d_i = 1 \\ 0 & \text{otherwise} \end{cases}$$

where

$$d_i = \begin{cases} 1 & \text{if } I_i > 0 \\ 0 & \text{otherwise} \end{cases}$$

and

$$I_i = h(w_i) + v_i$$

This means for each units there is an index I_i , function of the set of variable \mathbf{W} , for which participation occurs when it raises above zero and, only for treated units, an index t_i , function of the set of variable \mathbf{Z} , identifying the level of treatment.

The reason for adopting this structure represents the basis of the approach that will be presented further: it assumes to consider separately the selection process that determines the program participation from the further process that identifies the level of treatment. As seen in the previous chapter, this approach is not what is commonly followed by the literature: the works that deal with the continuous treatment setting generally focus on the specification of a model for the treatment level, given the set of pre-treatment variables, without explicitly consider the selection rule that identifies the non-treated and the treated units, at any level. Thus, there are no distinctions between all the different doses of treatment and the treatment at the level zero, that is no

treatment. Instead here, the idea is to specify first an assignment rule for the selectivity process that determines if a unit is treated or not, and subsequently a model for the specification of the level among the selected units. To justify this approach it might be noted it is reasonable that the selection rule and the treatment level assignment can be influenced by different variables: adopting different specifications for the two processes might be helpful for considering these different factors. Moreover, in this way, there is a distinction between the treatment at the level zero and any strictly positive dose of treatment. Then, the treatment interval \mathcal{T} is split in two parts: \mathcal{T}^0 which includes only the treatment level equal to zero. The other one, \mathcal{T}^+ , regards the set of positive treatment doses. In this way, the set of non participants is maintained separated from the set of participants through the full analysis, and this allows to identify the two different selection processes.

Actually, from a theoretical point of view, this approach could be embedded in the concept of a generalized propensity score, where the probability of a given output is estimated by a set of covariates and by a level of treatment, depending to the same covariates. However, our proposal has some empirical advantages: it exploits the full information set containing treated and not treated units; it contains a more efficient estimation of the two processes (the selection and the level of treatment); it can incorporate some empirical recognized restrictions on the relation between the two processes.

It is worth to note, however, that this distinction becomes meaningful if it is referred to two particular situations. The first one regards the parameter of interest one wants to estimate. In particular, when the parameter regards the effects of the levels of treatment with respect to the non treatment case, it might be important to consider the two selection process, as mentioned before. Instead, if one wants to estimate the effects of a level of treatment with respect to another level, the first selection process that specifies the treatment status might be not considered and the evaluation might be carried out only among the treated units. It might be the case, for example, of the estimation of the effects between units treated at “high” level and units treated at “low” level. On the other hand, this distinction becomes meaningful when one of the two components of the selection process, or both, is known (or partly known). In these situations, the evaluator might decide to use this information in order to better model and predict the selection assignment rules and the causal effects.

4.3 The parameters of interest

As in the binary treatment case, the possible counterfactual of interest, in a continuous setting, might be different: for example one might be interested in the comparison of the state of the world in the presence of an intervention at a particular level of treatment to the state of the world if the program did not exist at all, or if alternative levels were applied. A full evaluation should consider all outcomes of interest for all person, both in the current state and in all alternative state of interest. However, when the treatment is continuous this analysis is more difficult because the potential counterfactuals are infinite and another source of variability is introduced, given the continuous treatment variable.

To be more specific the treatment effects are now influenced by three components: the treatment level, the heterogeneity among the units and the stochastic component:

$$\alpha_i = f(T, i, \varepsilon)$$

where α_i represents the treatment effect for the i -th unit.

A complete evaluation analysis has to consider all these sources of variability. With respect to the binary treatment case, there is an additional component, that is the one induced by the treatment variable. Then, apart from the error term ε , the heterogeneity issue can be interpreted in two ways. First, the effects might be different among the levels, and that is what this work wants to focus on. Secondly, for each level, the effects may vary among units. That is the traditional heterogeneity problem in the literature of the program evaluation with binary treatment. Recent developments on this topic deal with the estimation of the distribution of the effects. Hence, the focus is not more on average treatment effects, but on other summarizes of the distribution, such as quantile estimations. This matter is not considered in this analysis, not because it is irrelevant, but rather because the focus is on the estimation of relation between effects and treatment levels.

Then, the statistical solution to the causality problem might be, as in the binary case, the transition from the individual to the group level counterfactual. Given the impossibility of observing the same person in different states at the same time, the focus might be on counterfactual means. This does not mean that the parameter of interest is an average treatment effect but rather an average effect among units evaluated at different treatment levels. The idea is to use the information of the treatment level to estimate different kind of treatment effects as a function of the doses.

4.3.1 Average treatment level effects

In order to study the relation between effects and treatment doses, what is proposed here is, first, an estimation of the average effect among units evaluated at different treatment levels. Assuming to have an infinite number of observations, or a series of finite number of observation for each treatment dose; a natural development of the treatment effects estimation in the continuous case should be the difference between the outcome of the units treated at each level with the outcome of the untreated units. That is the average treatment effect for the t -th level,

$$\alpha(T)^{ATE} = E[Y(T) - Y(0)]$$

for a person randomly drawn from the population, and the average treatment effect on the treated at the t -th level,

$$\alpha(T) = E[Y(T) - Y(0)|T = t]$$

for a person randomly drawn from the subpopulation of the participants at the level t . In both cases, the expected value is taken over all the observations treated at the same level. This latter parameter $\alpha(T)$ will be called the average treatment level effect (ATLE) and it will be the parameter this analysis will focus on. This may be justified as it happens in the binary case, where the average treatment effect on the treated (TTE) represents the parameter that had received more interest in the current literature (Heckman and Robb (1985) and Heckman *et al.* (1997b)). In fact, it is reasonable to argue that the subpopulation of treated units is often of more interest than the overall population in the context of narrowly targeted programs.

As in the binary case, only under the assumption of homogeneous treatment effect among units at each level $t \in \mathcal{T}$, all these parameters are identical, but it is obviously not true when treatment effects vary among individuals. However, as mentioned before, this heterogeneity issue is left for further studies and the focus here is on average treatment effect at each dose.

The framework presented above is obviously based on the potential outcome approach developed by Rubin (1974); it means that ex-post only the outcome corresponding to the program in which the individuals participate is observed. However, in this setting another “missing data” problem arises: because of the continuous nature of the treatment it would be difficult to cover all the possible levels and more to have different

units treated at the same level.

To go over this problem and to obtain an appropriate estimation of the counterfactuals of interest this work proposes to adopt a matching approach, with the aim of eliminating any biases associated with differences in the covariates by pairing similar units.

4.3.2 Treatment dose function

Once estimated the average treatment level effects for each observed treatment level, the next interesting object to estimate regards the specification of the relation between effects and levels

$$\alpha = f(t, \varepsilon)$$

in order to estimate the entire function of average treatment effects over all possible values of the treatment doses. The idea is to study if the treatment level differently influences the effects on the response variables.

To study this relation different approaches might be chosen, also with respect to different hypothesis and structures one wants to impose. What is proposed here is a parametric versus a non parametric approach.

In order to investigate if the treatment level differently affects the response variable, what we propose is to use an OLS estimator imposing a quadric relation between effects and level of subsidies:

$$\alpha = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon$$

A non linear estimation instead of a simple linear regression model is preferable in order to better detect some heterogeneity of the effects. A simple regression model would turn out the correlation between the two variables: it might be a primary information, but it implies a proportional structure of the treatment level and a curve of the effect that is a straight line. This might be a very strong assumption. Our idea is that the average impact can hide some relevant effects at some point of the subsidy distribution. Some considerations on this issue will be also discussed in section 4.3. For that reason we propose other specifications that can include some quadratic or higher order relations between effects and levels.

To model the treatment level effects, we also adopted another parametric approach: the quantile regression. The ability of quantile regression models (Koenker and Bassett (1978)) to characterize the heterogeneity impact of variables on different points of an

outcome distribution, makes them appealing for evaluating the effects of policy interventions. The quantile regression has been recently used in DID models for evaluating the effects of policy changes by Athey and Imbens (2006). The authors improve this technique to estimate the entire counterfactual distribution of outcomes that would have been experienced by the treatment group in the absence of the treatment and for the untreated group in the presence of the treatment. Restating our problem of evaluating the effects of policy interventions in a quantile regression framework allows us to investigate if treatment groups have been differently beneficiated from the treatment and to provide an analytical description of the effects distribution with respect to the treatment doses.

The relation between average effects and treatment level may be evaluated also by adopting a non parametric approach that allows to estimate this relation regardless of any functional form. What is proposed here is to use a non parametric mean regression estimator, the Nadayara-Watson kernel estimator (see Nadayara (1964)):

$$E[\alpha|T = t] = \frac{\sum_{i=1}^m \hat{\alpha}_i K(\frac{t-t_i}{h})}{\sum_{i=1}^m K(\frac{t-t_i}{h})}$$

where K is a kernel function, $\hat{\alpha}_i$ is the ATLE for the i -th level and m is the number of the estimated ATLE.

A graphical analysis of the estimated function might be useful to easily interpret the relation between levels and effects of treatment and to see if there are some relevant effects at some point of the subsidy distribution. However, it is worth to keep in mind that in the analysis of the relation between impacts and treatment level, average causal effects are computed by comparing treated against non treated units. That is, our method is able to estimate the causal effect due to the participation ruling out the differences between these two groups. This means that, in general, there could be some differences among treated units (at different levels). That is, participants might be different not only with respect to the level of treatment. Then, in order to evaluate the impact of the amount of granted subsidy on treatment effects one should be careful in the interpretation of this relation. With the proposed methods we are able to evaluate if there is some heterogeneity with respect to treatment level, but this different impacts can be interpreted as causal treatment effects only with respect to the non-treatment status.

Another important consideration, that cannot be underestimated when a program

consists of a continuous treatment, regards the selection process for the level of treatment. In particular there are two possible opposite alternatives: the case of a random versus a non random treatment level assignment. Let consider separately this two different situations.

4.4 The random treatment level case

This situation represents the simplest case: the selection process that identifies the level of treatment is random. This means that, given the first selection process that determines which units are treated and which are not, the next level assignment rule is random. Thus, it is assumed that, as in the binary case, the participation decision can be identified by the following rules:

$$I_i = h(w_i) + v_i$$

that means for each units there is an index I , function of the set of variable \mathbf{W} , for which participation occurs when it raises above zero. V is the error term and

$$d_i = \begin{cases} 1 & \text{if } I_i > 0 \\ 0 & \text{otherwise} \end{cases}$$

where, in this case, D is a dummy variable indicating the treatment status: in particular $D = 1$ if the individual has been treated, at any positive level, and $D = 0$ otherwise, that is if the unit is not treated. The next process identifying the treatment level is random: it means there is no more a function identifying the treatment level, but only a function that determines the treatment status, that is the index I . It is not needed anymore a continuous variable T indicating the treatment level, because the distinction among units is only between treated and non treated, and all the levels, bigger than zero, are equally likely. The treatment level assignment rule might be simply written as:

$$t_i = \begin{cases} RW & \text{if } d_i = 1 \\ 0 & \text{otherwise} \end{cases}$$

where RW stands for a random walk process. This represents the fundamental assumption of that particular case:

Assumption: *Random hypothesis:*

the treatment level variable T is allocated at random

that means the doses are given to all sample members with equal probability. If this assumption is valid it is implied, by design, that T will be independent of any other influences, whether observed or unobserved. Then, it implies independence between the levels and any observable variables. To be more specific, it can be written:

$$T \perp \mathbf{X} \quad (4.1)$$

where \mathbf{X} is the set of observable covariates.

The main advantages of this type of data come from the randomization process: controlling for the first selection assignment, treatment versus control, removes any biases associated with differences in the covariates determining this process. The further randomized process allows to rule out bias caused by self-selection among treated units, as individuals are randomly assigned to the treatment level of the program and there are no relations with other factors.

4.4.1 Estimation of the effects: a matching approach

Given the assumptions stated above, the estimation of the treatment effects might be not so different with respect to the binary treatment case. In particular, if the parameter of interest were the average treatment effect, it might be applied the methods used in the traditional evaluation policy with a binary treatment. Because there are no significant differences among the levels, it is sufficient to eliminate the bias associated with differences in the observable and unobservable components between the group of the treated and untreated units.

However, here the focus is on the ATLE parameter and on the curve of the treatment effects, as a function of the levels. For that reason, what should be done is to find some properly development of the traditional methods in order to estimate the new parameter of interest, or equivalently, to take in account for the continuous dimension.

What is proposed here is an estimation of the curve using a matching approach: the choice of this particular method comes from its nice features. In fact, it represents one of the most popular methods in recent works that deal with the program evaluation, because of its flexibility and its ability to simulate an experimental setting.

The idea is to use the properties of this method to eliminate the differences on the covariates between the treated and untreated units by combining individuals of the two

groups. Then, the mean difference between treated and control units at each observed level represents an estimation of the average treatment effect for that particular dose and an estimation of the entire curve of the effects might be obtained by modelling the estimated effects computed for each level.

In the random treatment level case, the assumptions the method relies on, are quite similar to the hypothesis needed for the binary case. Only two relevant treatment status are recognized: the non participants, that is $D = 0$, and the participants, that is $D = 1$, that means $t_i > 0$. Then, the main assumptions of this approach are:

Assumption 1: *Conditional independence*

$$Y(0) \perp D \mid \mathbf{X} \quad (4.2)$$

conditional on the set of observables \mathbf{X} , the non-treated outcomes are independent of the participation status.

Assumption 2: *Common support*

$$0 < P(d_i = 1) < 1$$

all treated individuals have a counterpart on the non-treated population and any-one constitutes a possible participant.

As in the binary case, the first assumption has the implication that for each treated observation, $Y_i(t_i)$, one can look for one or more non-treated observations $Y_i(0)$ with the same \mathbf{X} -realization and be certain that this $Y_i(0)$ constitutes the correct counterfactual. The second assumption ensures that each participant can be reproduced among the non-participants.

Let S represent the common support of \mathbf{X} . Then, under assumption 2, S is the whole domain of \mathbf{X} . The matching estimator proposed for the average treatment effect at the level t , ATLE, is the empirical counterpart of

$$\begin{aligned} \alpha(t) &= E[Y(T) - Y(0) \mid \mathbf{X} \in S, T = t] \\ &= \frac{\int_S E[Y(T) - Y(0) \mid \mathbf{X}, T = t] dF(\mathbf{X} \mid D = 1)}{\int_S dF(\mathbf{X} \mid D = 1)} \end{aligned}$$

This result represents the expected value of the program impact for the t -th level: it is the simple mean difference in outcomes over the common support S , weighted by the distribution of participants. We can consider all the treated units together, because of the random treatment level hypothesis. As before, the expected values is taken over observations with the same treatment level, assuming there are available information.

Then, the general form of the matching estimator is given by

$$\hat{\alpha}(t_i) = E \left[y_i(t_i) - \sum_{j \in C} a_{ij} y_j(0) \right] \quad (4.3)$$

where C represents the comparison group, a_{ij} is the weight placed on comparison observation j for individual i and t_i stands for the observed level of treatment for the i -th unit. That is, the estimator can be computed only at the observed treatment levels t_i . The expected value is taken among units treated at the same level t_i . Hence, it might happen that the average among treated units will disappear: it might be the case when it is impossible to empirically observe two individuals with the same treatment dose realization. This has some implications on the choice of the matching procedure to use: an accurate evaluation has to be done in order to properly consider the precision and the robustness of the estimation. It is clear there is a trade-off between these two aspects: considering separately each single treatment level $t(i)$ allows to increase the precision and the detail of the effects estimation that is evaluated at each level t of interest. On the other hand, this might decrease the significance and the robustness of the estimation, because the estimation of the treatment level effect is obtained only comparing one treated observation. That is why the choice of the matching procedure to use might become a very relevant issue for the analysis.

Propensity score matching

In order to improve the implementation of matching and to solve the dimensionality problem a feasible alternative is to match on a function of \mathbf{X} . Given the random treatment level assumption, the conditional independence assumption (Assumption 1) remains valid if controlling for the propensity score $p(x) = P(D_i = 1 | \mathbf{X}_i)$, that is the propensity to participate given the set of characteristics \mathbf{X} , instead of \mathbf{X} :

$$Y(0) \perp D | p(x).$$

As mentioned before, the relevant assignment rule is only the one that defines the treatment versus the non treatment status: once obtained the treatment, the levels are independent of any variables and the level effects can be simply identified by the difference between participants and non participants.

As in the binary case, when using $p(x)$, the comparison group for each treated individual is chosen with a pre-defined criteria of proximity between the propensity scores for each treated and the controls. The next step is that of choosing the appropriate weights to associate the selected set of non treated observations. The possibilities are the same as in the binary case: nearest neighbor matching, kernel matching, radius matching, etc... Different weighting schemes define different estimators, but the form of the matching estimators is again given by $\hat{\alpha}(t_i)$ in (4.3).

The dimensionality problem, the choice of the measure of distance and of the matching algorithm are discussed deeper in the section 4.7, considering also the more general case of non random treatment level selection.

4.4.2 Test for the independence assumption

The estimator presented above relies on the random hypothesis on the treatment level variable. Then, a relevant issue regards a way to verify its validity. If the assumption is satisfied, it is implied, that T will be independent of any observable variables X . That is, observations that receive any positive level of treatment must have the same distribution of observable (and unobservable) characteristics. This is what is commonly called covariate balance. In other words, treated units should be on average observationally identical.

There are many issues involved with choosing appropriate tests. However, the focus here is not on this issue and what is proposed here has not anything to add to this vast literature.

In order to verify the independence assumption two solutions are proposed: the first one tests that the means of each characteristic do not differ between treated units at different levels. The second solution tests the absence of correlation between each characteristic and the treatment variable.

The first solution might be implemented computing the following steps:

1. split the sample in k equally spaced intervals of the treatment level, for $t > 0$.
2. for each characteristic test that the means do not differ among the different intervals.

3. if the means of one or more characteristics differ, the independence assumption is not satisfied.

As regards the first point, the number of intervals k has to be determined by the user after an accurate evaluation on the size and on the support of the sample that is available. A possible solution might be the one of repeating the tests for different partitions of the full interval \mathcal{T}^+ . If the treatment level assignment is really random any partition should bring to not reject the hypothesis that the means of each characteristic are equal among the intervals.

To test the null hypothesis

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_j = \dots = \mu_k$$

where μ_j , is the mean of a characteristic X for the j -th intervals, versus the alternative hypothesis that at least one of the equality does not hold, a simple analysis of variance can be computed, in order to test for differences in means. This test must be repeated for each interval for the treatment level. To be more complete, because it should be important not simply to test for differences of means some test of the equality of distributions, or some graphical summaries such as quantile-quantile plots that compare the empirical distribution of each covariate, might be performed.

In any case, if the null hypothesis is rejected, for one or more variable X , it follows that, for these characteristics, at least one mean is statistically different from the others, computed for each interval. This implies that units treated at different levels are not on average observationally identical and that the independence assumption between the treatment variable T and the set of observable characteristics \mathbf{X} is violated.

Another solution to test the independence assumption is to specify a relation between the treatment variable T and the observable regressor X in order to verify the absence of correlation between the variables. This solution might be implemented computing the following steps:

1. for each X regress the treatment variable T on this covariate X
2. test the coefficient of the regression
3. if at least one coefficient is different from zero, the independence assumption does not hold
4. otherwise specify more complicated models, such as non linear models and test again the coefficient of the covariate X

As regards the first point, the linear regression between T and each X can be written as:

$$T = \beta_0 + \beta_1 X + \varepsilon$$

Then, the null hypothesis to test is $H_0 : \beta_1 = 0$. It can be computed a simple t -statistic for the coefficient β_1 . If this hypothesis is rejected for at least one covariate, it is implied that a linear relation exists between the treatment level and this regressor. This means the independence assumption does not hold and the treatment level assignment is not random. On the other side, if the null hypothesis is not rejected for all covariates, it does not necessary mean that the independence assumption holds, but only that the treatment variable is not linearly dependent from the covariate X . That is why what is suggested is to specify other models that contains higher order relations or interactions between different regressors X . For each specification, just test the statistical significance of the coefficient of the X .

Whatever solution is chosen to verify the random hypothesis, if the results reveal that it might not hold, the estimator presented above could not be more appropriate, because it does not consider the differences among units treated at different levels. This means that an accurate evaluation of the random hypothesis on the treatment level assignment rule might constitute a starting point for a complete analysis on the treatment effects in a case of continuous doses. Once tested this assumption might not hold, it has to be taken into account that there should be some bias associated with differences in the treatment level. The next section will deal with this issue and will extend the identification and estimation of causal treatment effects in the case of non random level assignment.

4.5 The non-random treatment level case

As mentioned before, in a case of no random treatment level assignment the selection process might be split in two parts. The first process, that is the participation decision, might be represented, as in the previous case, by the following rule:

$$I_i = h(w_i) + v_i$$

where \mathbf{W} is again a set of variable, V is the error term and

$$d_i = \begin{cases} 1 & \text{if } I_i > 0 \\ 0 & \text{otherwise} \end{cases}$$

where D is the dummy variable indicating the treatment status. Given a positive value for the index function I , the next process deals with the identification of treatment levels that are not more equally likely among treated units. Selection into the different levels of treatment, among participants, might be represented by:

$$t_i = g(Z_i) + u_i$$

where T is the treatment level variable, \mathbf{Z} is a set of variables and U is an error term.

As specified before, the full selection process might be written as:

$$t_i = \begin{cases} g(Z_i) + u_i & \text{if } d_i = 1 \\ 0 & \text{otherwise} \end{cases}$$

where

$$d_i = \begin{cases} 1 & \text{if } I_i > 0 \\ 0 & \text{otherwise} \end{cases}$$

and

$$I_i = h(w_i) + v_i$$

This framework follows the traditional literature on selection into treatment in the program evaluation setting: it implies a linear additively separable relation between the sets of covariates \mathbf{W} and \mathbf{Z} and the error terms V and U respectively. The main advantage of this framework is that the sets \mathbf{W} and \mathbf{Z} can contain different variables: this would mean allowing different factors to influence the two processes. This might be not so rare in practise.

It is worth to remember that the first process involves all individuals, participants and not, while the level assignment rule will be specified only for treated units, that is what the previous literature on this topic concentrates on.

4.6 A new approach: the use of a matching procedure

Among methods used in literature to assess public policy interventions the one that recently has received more attention is the matching method. As before in the random level case, the main idea here, is to develop this traditional method in order to take into account the treatment level variable and to estimate the new parameter of interest.

The proposed estimators will estimate the ATLE, considering the selection process issue stated above, in two different ways:

- structured form of the selection process;
- non-structured form of the selection process.

The first case represents a situation where the selection process is partly known. To be more precise, in this case the evaluator has some information on the selection process. It might refer to one of the two components of the selection rule (participation decision and treatment level assignment), or both. It might be the case, for example, of a selection rule determined by some threshold of a ranking, or of a situation where some constraints on variables are imposed by some authority. This does not happen in the second situation, where the selection process is a function only of the individuals' choices and the evaluator has not further information governing the assignment to treatment. If we are in a situation where some aspects of the selection criteria are known, why does not use them and try to estimate and predict the participation rule in a more detail and efficient way?

In particular, this distinction might be relevant when a matching approach is used. Consider, for example, the case where some constraints on the treatment level assignment are imposed: using a unique function of the observable variables summarizing the selection processes, even if an ignorability assumption is satisfied, might bring into a situation where some treated units are paired with control units with different values on the variables governing the treatment level assignment, that might be impossible given the imposed restrictions. In this case, what might be done is to better specify the two selection processes in order to reduce these “unlikely” matchings.

These two cases will be kept separated throughout the chapter and the results of their implementation will be compared in the last chapter, where the empirical application will be presented.

4.6.1 Structured form of the selection process

The idea is to split the selection process in order to better predict it and to take into account relations between these two processes, remembering also that some covariates might affect each process, but with a different weight. The idea is to consider these differences in order to avoid matching between units with similar values in the variables that specify one process but different values in the others. The first step will identify the participation decision and units will be matched on the basis of similar value of a set of covariates. Among matched units in the first step, another matching procedure will pair units with similar value in the second process that identifies the treatment level. Let consider separately the two matching procedures.

The participation decision

The participation decision process and the subsequent matching procedure might be considered not so different with respect to the random level case: the aim is again to eliminate the bias associated with differences in the observable and unobservable components between the group of the treated and untreated units by combining individuals of the two groups.

The main assumptions this procedure relies on are equal to the hypothesis needed for the random level case. This comes from the fact that the selection process is split in two components and this first one is identical to the previous case. Only two relevant treatment status are considered: the non participants, that is $D = 0$, and the participants, that is $D = 1$, that means $t_i > 0$. Then, the main assumptions of this approach are the same as before:

Assumption 1: *Conditional independence*

$$Y(0) \perp D \mid \mathbf{W}$$

Assumption 2: *Common support*

$$0 < P(d_i = 1) < 1$$

As before, the second assumption ensures each participant can be reproduced among the non-participants. The first assumption has again the implication that non-treated observations $Y_i(0)$ paired with each treated observation $Y_i(t_i)$, with the same \mathbf{W} -realization, constitutes the correct counterfactual. However, there is a fundamental

difference with respect to the previous case: for each treated observation $Y_i(t_i)$, it is preferable that one looks for more than one non-treated observations $Y_i(0)$ with the same \mathbf{W} -realization. This is necessary, because the matching solution proposed is a two-steps procedure. To be more precise, it is a sequential matching: for each treated observation $Y_i(t_i)$, the set of non-treated observations $Y_i(0)$ paired in this first step will constitute the non treated group to be used in the second step of the matching. This second step will match units with a similar Z -realization, where the set Z describes the treatment variable T . If in the first matching a one-to-one method is chosen, that is one looks only for one non-treated observation, the next control group for the further step, for each treated units, will be constituted only of one non-treated observation. Then, if there are no similar control units to match with, these treated units have to be excluded from the analysis. Instead, if one looks for more than one control observation for each treated units, it will be possible to compute the second matching, among the units chosen in the first step.

This obviously has some implications in the choice of the algorithm that assigns units to matched sets: what has to be used is a method that pairs more than one control unit to a single treated unit. However, this is not a strong limitation: in literature there is a wide choice among this kind of procedures. The most popular ones are the nearest neighbor one-to- k algorithm, the radius matching, the full matching and the kernel matching. However, each solution has to be properly evaluated with respect to the contest the analysis refers to.

To be more specific, let consider the nearest neighbor one-to- k matching: by this algorithm, for each treated unit are selected the k most similar units among the group of the non participants. Imposing the number k of controls to be matched might bring to use control observations that are not so similar from the treated paired unit, especially if k is big with respect to the available sample size. A possible solution to this problem can be to use a full matching algorithm: by this method each matched set consists of either a treated unit with at least one control or a control unit with one or more treated units. Obviously, the first proposal is the one of interest in this work. The main advantage of this algorithm is that it is not required every treated unit to have the same number of control. On the other hand, there could be some treated units with few or even only one control matched unit. This constitutes the main drawback of that procedure: in order to apply the two steps matching algorithm proposed by this work it is necessary that treated units matched at the first step have more than one paired control units, otherwise these units have to be excluded from the analysis.

The radius matching might be a valid proposal: this algorithm admits for each

treated unit to be matched with more than one excluded unit. In particular, the matches are made only if the distance between treated and control unit is lower a predetermined level. The main disadvantage of this procedure is that the treated unit is bypassed and no match is made to this individual if there are no excluded units within the fixed radius. Moreover there could be the case again of few paired control units for each participant. However, different size of the radius might be chosen and compared.

The solution proposed by the kernel matching might be in principle the more suitable for this work: by this algorithm the entire comparison sample is used. It is a smooth method that reuses and weights the comparison group sample differently for each treated unit with different covariates. The weight given to non treated unit is in proportion to the closeness between excluded and treated unit. The main advantage of this method applied to the proposed analysis is that there is no lost information because the entire control sample is used either in the first and in the second step matching. However, this algorithm might be not so suitable in this application: we do not want to pair all units, because some matching would be impossible, given the imposed restrictions of the selection process.

These brief considerations about the choice of the algorithm that identifies the most similar comparison units reveals that an accurate evaluation has to be carried out in order to consider all the limits and advantages of each method, given the available information. It is clear that there are some trade-off that should guide the choice. For example in the nearest-neighbor one-to- k matching there is a trade-off between precision and the number k of control units to match: the higher is k , the higher will be the size of the control sample for the second step matching. On the other hand an high value of k might raise the precision of the matching in term of closeness between units. For the radius matching a serious problem might regard the choice of the width of the radius. Thus, all these aspects have to be considered and possibly different solutions may be computed and compared.

Independently from the algorithm, it is important to remember that the control observations matched with each participant will constitute the control sample for this unit in the further step of the matching. This implies a different control sample for each treated individual i in the second step procedure and an accurate evaluation on the weights to attribute at each paired unit. The new control sample selected for the i -th unit, will be denoted by M_i , while $\mathbf{M} = \cup M_i$ denotes the full sub-sample of selected non participants.

The treatment level assignment

The treatment level assignment process in the case of non random levels might induce some bias associated with differences in the observable and unobservable components between treated units at different levels. The focus of an accurate analysis is to rule out these sources of bias, given the available information. Also if there are some imposed restrictions, selection into different treatment levels is in part influenced and determined by treated individuals. This kind of information are often not available from the researcher that conducts the analysis and this might bring some bias in the estimation of the effects.

The aim is again to eliminate this source of bias following a matching approach, on the basis of the matching procedure applied in the previous step between treated and non treated units. The idea is again to constructs the correct sample counterpart for the missing information on the treated outcome at different levels had they not been treated by paring each participant with member of non-treated group. However, in this case, the control group will be composed by the non participants selected with the first selection process, \mathbf{M} . Then, the key assumption is that, conditional on a set of observables \mathbf{Z} , treated at the level t and non-treated individuals are comparable in what respect to the outcome in the non-treatment case. This implies an independence assumption between the outcome of the selected non participants and the level of treatment conditional on the set \mathbf{Z} . To be more precise, the main assumption is:

Assumption 3: *Conditional independence*

$$Y(0)' \perp T \mid \mathbf{Z}$$

where $Y(0)'$ is the outcome of the non-treated units that belong to the sub-sample \mathbf{M} , that are the most similar in terms of the variables \mathbf{W} identifying the first selection process. This means it is no necessary to specify again the independence assumption of the outcome variable and the treatment conditional on the set of variable \mathbf{W} , because it follows directly from assumption 1 (4.2). Moreover, it is not required independence of the outcomes of the entire control sample C , but only on the selected sub-sample. In some sense, it may be considered as a weak form of independence if compared to the stronger assumption of independence between the outcome of the full set of non-treated outcomes $Y(0)$. However, if the matching algorithm chosen in the first step uses the entire comparison sample C , such as the kernel matching, the two assumption are equivalent. It is not necessary to assume independence between the treatment level and

the potential outcomes for the treated individuals $Y(T)$ if the parameter of interest is the ATLE.

Assumption 3 implies that it can be written:

$$\begin{aligned} E[Y(0)'] &= E_Z[E[Y(0)'|Z]] \\ &= E_Z[E[Y(0)'|T = 0, Z]] \\ &= E_Z[E[Y(0)'|T = t > 0, Z]] \end{aligned} \tag{4.4}$$

where the independence assumption is used in the second and third equality. Hence, it is possible to express the counterfactual of interest as a function of the observed data.

Although the conditional independence assumption identifies the conditional potential outcome $E[Y(0)'|T = t > 0, Z]$ through observation of non participants, this identification holds only for all z for which there is a positive probability that participants at the level t are observed with characteristic z . That is, we need to ensure that each treated observation can be reproduced among the non-treated. The second matching assumption is therefore a common support assumption. In particular in this case, it is necessary that all participants treated at the level t have a counterpart on the population of the non treated units M_i selected by the first matching algorithm. Then, the common support assumption can be written as

Assumption 4: *Common support*

$$0 < P(t_i = t|Z_i) < 1, \quad \forall t \in \mathcal{T}^+$$

with $i \in M_i$ and $P(t_i = t|Z_i)$ is the probability that an individual with characteristics z is treated at the level t .

Assumptions 3 and 4 allow to apply the second step of the matching procedure proposed and to estimate the ATLE: participants and non participants are now comparable in the sense that the only difference between the two groups is program participation.

The two-steps matching estimator

Before proceeding with the definition of the matching estimator, we would like to clearly state how the parameter of interest ATLE is identified, given the assumptions discussed above.

The first set of assumptions, relative to the participation decision process, allow to identify and to estimate the first counterfactual of interest $Y(0)'$, that is the set of

outcomes of the non-treated units if they were treated. Let S represent the common support of \mathbf{W} , that is the subspace of the distribution of \mathbf{W} that is both represented among the treated and the control groups, assumption 2 implies that S is the whole domain of \mathbf{W} . Together with assumption 1 it can be identified the set of potential conditional outcomes by:

$$(Y(0)|D = 1, W \in S) = (Y(0)|D = 0, W \in S) = (Y(0)|W \in S)$$

The second term of the equality can be estimated from observed data by the outcome values of the non-treated units adjusted for the values of the \mathbf{W} variables. For each treated unit i may be found the most similar set of non treated unit in term of \mathbf{W} in order to estimate the correct counterfactual for each participant:

$$Y_i(\hat{0}) = \sum_{j \in C} w_{ij} Y_j(0)$$

where C represents the comparison group and w_{ij} is the weight placed on comparison observation j for individual i . Then, the estimator for $Y(0)'$ is simply given by:

$$Y(\hat{0})' = \bigcup_{i \in T} Y_i(\hat{0})$$

where T represents the treatment group.

The second set of assumptions, which regards the treatment level assignment, allows to identify the parameter of interest ATLE. In particular, given assumptions 3 and 4 together with the equation (4.4) it can be written:

$$\begin{aligned} E[Y(T) - Y(0)'|T = t] &= E[Y(T)|T = t] - E[Y(0)'|T = t] \\ &= E[Y(T)|T = t] - E_z\{E[Y(0)'|Z, T = t]|T = t\} \\ &= E[Y(T)|T = t] - E_z\{E[Y(0)'|Z, T = 0]|T = t\} \\ &= E[Y(T)|T = t] - \int E[Y(0)'|Z, T = 0] f_{Z|T=t}(z) dz \end{aligned} \tag{4.5}$$

where $f_{Z|T=t}(z)$ denotes the density of \mathbf{Z} among the participants at the level t . The former term is identified by the sample mean outcome of the participants at the t dose, and the latter term can be estimated by adjusting the average outcomes in the no treatment group for the distribution of \mathbf{Z} among participants at the level t and by

replacing $Y(0)'$ with its estimator $Y(\hat{0})'$ identified by the first matching estimator.

Although the conditional independence assumption identifies the conditional potential outcomes $E[Y(T|Z)]$ through observations on participants at the level t , this identification holds only for all z for which there is a positive probability that treated units at the dose t are observed with characteristic z . The common support assumption deals with this issue: let S_t represents the common support of \mathbf{Z} for the t -th level, that is, the subspace of the distribution of \mathbf{Z} that is both represented among the treated at t and the control groups. Under assumption 4, S_t is the whole domain of \mathbf{Z} for all t in \mathcal{T} .

After a proper identification strategy has been established, the parameter of interest can be estimated. A class of matching estimators of equation (4.5) is obtained by replacing the distribution $f_{Z|T=t}$ by its empirical distribution function in the sample of treated units and estimating nonparametrically the conditional expectation function $E[Y(0)|Z = z, T = 0]$ from the non participants sample. Then, a general form for the proposed two-step matching estimator is given by:

$$\hat{\alpha}(t_i) = E \left[y_i(t_i) - \sum_{j \in C} m_{ij} y_j(0) \right] \quad (4.6)$$

where C represents again the comparison group, the expected value is among units at the same treatment level t_i and m_{ij} is the new weight placed on comparison observations j for individual i that comes from the two matching procedures. In particular,

$$m_{ij} = \frac{w_{ij}^1 w_{ij}^2}{\sum_{j \in C} w_{ij}^1 w_{ij}^2} \quad (4.7)$$

where w_{ij}^1 and w_{ij}^2 are the weight placed on comparison observations j for individual i in the first and second matching respectively. It is easy to understand that if one express the outcome variable as first difference, before and after the program, this matching estimator may be extended into a matching diff-in-diff estimator, as it will be in the empirical application.

4.6.2 Non-structured form of the selection process

Where there are no information about some aspects of the selection process or any particular restrictions or constrains are imposed to some selection variables, what might

be done is to follow that part of the literature that deals with continuous treatment adopting the so called generalized propensity score approach (Hirano and Imbens (2004) and Imai and van Dyk (2004)). The idea is to use a unique function of the observable covariates $\mathbf{X} = (\mathbf{W}, \mathbf{Z})$. If an appropriate specification for the propensity score is found, such that the balancing property is satisfied, and given a strong ignorability assumption, it might be sufficient to pair units with respect to the values of this propensity function. The differences with the strategies proposed in literature mainly refer to the parameters of interest. Here the concern is on the differences between treated and non treated units, that is on the comparison of some outcome variables in the presence of an intervention (at a particular level of treatment) to the same variables if the program did not exists at all, and not if alternative levels of treatment were applied.

To be more precise the underlying assumptions and the matching estimator will be clearly defined. However, it is a case not so different from a binary treatment setting; the main difference is that here, the set of observable covariates is composed of a higher number of variables. It is worth to note that the sets \mathbf{W} and \mathbf{Z} might have some common variables. It is clear that the set \mathbf{X} will contain all these common variables together with the specific covariates of each set.

Then, the main assumptions of this approach might be written as:

Assumption 1: *Conditional independence*

$$Y(0) \perp D, T \mid \mathbf{X}$$

Assumption 2: *Common support*

$$0 < P(d_i = 1) < 1$$

These assumptions imply the existence of a set of variables such that, conditioning on them, the counterfactual outcome distribution of the participants is the same as the observed outcome distribution of the non-participant. Once tested this balancing property, it is possible to find for each treated observation Y^1 a non treated (set of) observation(s) Y^0 with the same \mathbf{X} realization.

As in a binary treatment framework, given the common support, to obtain a measure of the treatment effect on the treated, individual gains from the program among the subset of participants who are sampled and for whom one can find a comparable non-participant, must be integrated over the distribution of observables among treated units and re-scaled by the measure of the common support. The only difference with

respect to the binary case is that the effects have to be identified and estimated at each treatment level value for which one is interested in. Thus, the matching estimator for the ATLE at the t -th level is the empirical counterpart of

$$\begin{aligned}\alpha(t) &= E[Y(T) - Y(0)|\mathbf{X} \in S, T = t] \\ &= \frac{\int_S E[Y(T) - Y(0)|\mathbf{X}, T = t] dF(\mathbf{X}|D = 1)}{\int_S dF(\mathbf{X}|D = 1)}\end{aligned}$$

where S^* is the common support. This result represents the expected value of the program impact at the treatment level t : it is the simple mean difference in outcomes over the common support S^* , weighted by the distribution of participants.

The matching estimator

In practise, to construct matches, it is needed again a measure of distance between the units with respect to the \mathbf{X} , in order to define the units in the comparison sample who are neighbors to each treated units i . Then, the general form of the matching estimator is given by

$$\hat{\alpha}(t_i) = E \left[y_i(t_i) - \sum_{j \in C} a_{ij} y_j(0) \right] \quad (4.8)$$

where C represents the comparison group, a_{ij} is the weight placed on comparison observation j for individual i and t_i stands for the observed level of treatment for the i -th unit.

4.7 Reduction of the multidimensionality

No considerations has been done yet on the distance measure to use in the matching procedures. This is not an irrelevant issue, especially if the set of covariates has a large dimension. This is what is commonly known as the dimensionality problem. Moreover, it is worth to note that these considerations might be applied both in the case of non-structured and structured form of the selection process. In particular, in the simpler first case the multidimensional set to consider regards the variables \mathbf{X} determining the two selection processes together. In the case of a more structured specification form for the selection process the dimensionality problem regards the two set of covariates \mathbf{W} and \mathbf{Z} .

The most common method of multivariate matching is based on Mahalanobis distance: it simply involves the estimation of the variance and covariance matrix of the set of regressors but it has serious limitation when the number of regressor is large. If the set of covariates consists of more than one continuous variable, multivariate matching estimates contain a bias term which does not asymptotically go to zero at \sqrt{n} . (Abadie and Imbens (2006)). A more feasible alternative is to match on a function of the sets of regressors to solve the problem of an high dimension vector of covariates.

4.7.1 Structured selection process

In this case, let consider separately the two selection processes and therefore the two sets of regressors. In the first step of the proposed procedure, that deals with the participation decision process, Rosenbaum and Rubin (1983) show that the conditional independence assumption (Assumption 1) remains valid if controlling for the propensity score $p(w) = P(D_i = 1|W_i)$ instead of \mathbf{W} :

$$Y(0) \perp D | p(W).$$

The matters regarding the choice of a pre-defined criteria of proximity between the propensity scores for each treated and the controls and of the weights to associate for each participant have been discussed in the previous section 4.6.1 relative to the participation decision. The main concern regards the choice of an algorithm that assigns more than one control unit to a single treated unit because of the sequential structure of the proposed matching procedure.

As regards the second selection process about the treatment level assignment, the idea is to extend and generalize the propensity score method applying it to a continuous treatment regime in order to reduce the dimension of the covariates set \mathbf{Z} on which the matching algorithm is carried out. Following the work of Imai and van Dyk (2004) what is proposed here is to find an appropriate (“propensity”) function of the variables \mathbf{Z} to model the treatment variable T . As stated by Imai and van Dyk, the propensity function is the conditional probability of the treatment given observed covariates \mathbf{Z} . Let $\theta(\mathbf{Z}) = E(T|\mathbf{Z})$ the parameter that uniquely represents the propensity function they show that, given the propensity function, the conditional distribution of the actual treatment does not depend on observed covariates, that is the balancing score property, and that the conditional independence assumption holds with \mathbf{Z} replaced by the propensity function. The parameter θ completely determines the propensity function. Then, matching on

the propensity function can be easily accomplished by matching on θ , regardless of the dimension of \mathbf{Z} . That is:

$$Y(0)' \perp T \mid \theta(Z).$$

The difference with the matching strategy proposed by Lu *et al.* (2001) regards the groups on which the pairs are accomplished. In particular, while Lu *et al.* (2001) suggest matching pairs among the group of treated units on the basis of similar values of θ but dissimilar values of the treatment dose, here the pairs are performed among the group of treated versus non treated group. Hence, the distance measure to use does not need anymore to consider for the levels. This is consistent with the idea of Lu *et al.* (2001): the pairs are computed among units with different treatment doses, that is treatment versus no treatment.

As in the binary case, given an univariate function of the observable variables, the comparison group for each treated individual is chosen with a pre-defined criteria of proximity between values of these functions for each treated and the controls. This criteria will also define the appropriate weights to associate the selected set of non treated observations for each participant. The possibility are the same as in the binary case: nearest neighbor matching, kernel matching, radius matching, etc..., with the difference that here the matching is performed two times. More details are given above in section 4.6.1. Again, different weighting schemes define different estimators, but the form of the matching estimators is again given by $\hat{\alpha}(t_i)$ in (4.6).

Relation between the two processes

It is not so unlikely to think that the two selection processes are in some sense related. In particular, it is reasonable to argue that the treatment level assignment might depend on the first participation decision rule. Consider, for example, the case of capital subsidies to firms, that represents the empirical application of next chapters: given the auction mechanism, higher is the probability of receiving the subsidy, lower is the amount of requested subsidies. On the contrary, there could be cases of positive relation between the two processes: higher the probability of being treated, higher the level.

To better identify and predict the selection process this relations might be taken into consideration and this could be done in different ways. What is proposed here is to condition the treatment level assignment to the participation decision rule. Thus, the function of the observable variables \mathbf{Z} , used to model the treatment variable T , is

conditioned to the propensity score. To be precise,

$$\theta(\mathbf{Z}) = E(T|\mathbf{Z}, p(W))$$

becomes the new parameter that uniquely represents the propensity function modelling the treatment variable T .

4.7.2 Non-structured selection process

This case is similar to the binary treatment setting: to improve the implementation of matching and to solve the dimensionality problem, a feasible alternative is to match on a function of \mathbf{X} . The conditional independence assumption remains valid if controlling for the propensity score $p(x) = P(D_i = 1|\mathbf{X}_i)$.

As in the standard matching, when using $p(x)$, the comparison group for each treated individual is chosen with a pre-defined criteria of proximity between the propensity scores for each treated and the controls. Once defined the neighborhood for each units the second choice regards the appropriate weights to associate each selected individual of the control group with the treated one. The several solutions are again the same discussed above: from weight equal to one to the nearest observation and zero for the others, to equal weights to all, otherwise, kernel weights.

Chapter 5

The law 488/1992 in Italy

5.1 Introduction

The purpose of this chapter and the next one is to provide a statistically robust evaluation of the impact of subsidies allocated by law 488/92 in Italy on subsidized firm, using the proposed nonparametric approach. We evaluate how the receipt of different level of financial assistance from public funds actually makes a difference to firm performance in terms of investment, new employment, profit and labor productivity.

In particular, in the first part of the chapter a detailed description of the law and of its selection mechanism will be presented, while the rest of the chapter will deal with the description of the data.

There are few studies concerning ex-post evaluation of the impact of L488. Furthermore, in the estimation of the counterfactual all the previous papers consider the 488 as a case of a binary treatment, and do not exploit the richness of the data set that includes information on the level of the subsidy by firm. We use the paper of Pellegrini and Centra (2006) to confront the results in the case of binary treatment approach with respect to the case of continuous treatment approach.

5.2 Capital subsidies in Italy

State aids to manufacturing and to service sectors, in the form of grants and subsidies, have been for many years a key component of regional policy in less developed Italian areas, such as Mezzogiorno. The use of such policy instruments has been aimed at

influencing the regional allocation of investments and employment, in order to increase competitiveness, self-sustaining growth, and new employment in low income regions.

Since post-war period, several policy instruments have been implemented to overcome the gap between southern regions and the most developed areas located in North-Center of Italy. Among the different policy instruments, the capital subsidies have been the most employed, either in the form of grants or reductions in the cost of borrowing. In the last two decades they became the core of the regional policy for the South. The rationale for such interventions has to be found on the need to fill the gap in the availability of private capital and, consequently, to increase productivity and production capacity. Subsidies were perceived as a “compensation” for the productivity gap of the disadvantages areas, due to the limited availability of public infrastructure and negative territorial externalities. As a consequence, the subsidized projects have never been selected on the base of economic parameters and criteria: all investing firms could participate and benefit, the only constrain was the availability of financial resources. This approach has maximized the number of subsidized investment projects, but it also has negative effects on policy additionality and efficiency Pellegrini (1999).

The end of the “extraordinary intervention”, the reduction in the financial resources, and the new approach of regional policy, more oriented to efficiency, competitiveness and self-sustaining growth imposed a radical change in the policy instrument to offer financial support to private investments in Mezzogiorno. This long phase of uncertainty ended with the design and the implementation of a new regional policy, named “new programming” (Barca and Pellegrini (2002)). In this policy framework a new instrument to subsidize investment in the so-called disadvantaged areas has been implemented: the L488/1992, founding private capital accumulation by project-related capital grants. The law allocates subsidies through a “rationing” system based on an auction mechanism which guarantees compatibility of demand and supply of incentives.

L488/92 has been in the last ten years the main policy instrument to encourage private investments. From 1996 (the first operative year) through 2005, this law has sustained about 40,000 projects with over 20 billions of Euros of subsidies, whereas investments have added up to over 70 billions of Euros, 70% of which in the South. The expected additional employment from these investments amounts to about 560,000 new units. After ten operative years, and in view of the extent of spending on L488, it is reasonable to investigate if the law made a difference (or not) to the industrial structure of Mezzogiorno in term of growth, employment and productive efficiency.

The literature aiming at an evaluation of the impact of subsidies on firm behavior is now extensive. It is generally accepted that regional capital incentives induce additional

investment (Faini and Schiantarelli (1987), Harris (1991), Daly *et al.* (1993) and Schalk and Untied (2002)), even if they can have unpleasant effects on income inequality across different areas (Dupont and Martin (2003)). Besides, they have some effect in attracting plants to low income areas (Faini and Schiantarelli (1987) and Midelfart-Knarvik and Overman (2002)). The employment impact of capital subsidies is more doubtful: the question is if the size of substitution effect, associated to the reduction in the user cost of capital relative to the labor cost, is larger or smaller than the output effect, related to the increase in production (and therefore in local labor demand), due to the reduction in total costs and to the attraction of new investment in the area (Schalk and Untied (2002)). Several studies found that the substitution effect outweighs the output effect (Driehuis and van den Noord (1998), Harris (1991) and Gabe and Kraybill (2002)), others found the opposite (Wren and Waterson (1991), Daly *et al.* (1993), Schalk and Untied (2002) and Rooper and Hewit-Dundas (2001)). Few studies evaluate if the receipt of financial assistance from public funds actually makes a difference to firm performance in terms of improved plant efficiency or productivity. Increase in investment, both in additional productive capacity and in replacement investment, modernizes the firm stock of equipment and results in higher efficiency and productivity. On the other side, capital subsidies can also have potential negative effects on productivity, by increasing allocation inefficiencies if lower relative capital costs leads firms to overinvest in capital, and by encouraging rent-seeking behavior by firms for subsidies (Harris and Trainor (2005)). All the studies show that the effects of subsidies on efficiency and productivity are negligible or negative (Lee (1996), Bergstrom (1998) and Harris and Trainor (2005)).

There are few studies concerning ex-post evaluation of the impact of L488. A positive effect of L488 on investment is found in Ministero dell'Industria (2000) and in Bronzini and de Blasio (2006). Carlucci and Pellegrini (2003) and Carlucci and Pellegrini (2005) present empirical evidence of a positive employment effect. Pellegrini and Centra (2006) also on turnover but not on productivity. Bronzini and de Blasio (2006) indicate the presence of (moderate) inter-temporal substitution: financed firms slow-down significantly their investment activity in the years following the program. In the estimation of the counterfactual all the previous papers exploit the auction mechanism that 488 uses to allocate the subsidies across firms. The group of subsidized firms are compared with the group of firms that applied for the incentives but were non financed, since they score low in the ranking. These non financed firms, then, are especially eligible to be part of a control group, as they show a propensity for investment and a need

to invest which is very similar to that of subsidized ones. As suggested by different authors (Brown M. A. and Elliott (1995), Carlucci and Pellegrini (2003) and Bronzini and de Blasio (2006)), the rejected application group is very similar to the treatment group in terms of its characteristics, and allows to isolate the effects of policy intervention. Moreover, all the papers consider the 488 as a case of a binary treatment, and do not exploit the richness of the data set that includes information on the level of the subsidy by firm. Therefore, in this work, we also check if the previous results are confirmed using a continuous treatment estimator. We will use the paper of Pellegrini and Centra (2006) to confront the results in the case of binary treatment approach. Moreover, the application includes the estimation of the entire function of average treatment effects overall possible values of the treatment levels. This function is estimated by adopting parametric and non parametric methods, using a flexible functional form. Therefore we can determine the fraction of treatment effect heterogeneity that can be attributed to different levels of treatment.

5.3 The Law 488/92

The law operates in the Mezzogiorno and in all so-called disadvantaged areas; these areas are either designed as Objective 1, 2 or 5b for the purpose of EU Structural Funds or subject to exemptions from the ban on state subsidies. The Law488 auctions are run on a yearly basis and take the form of a project-related capital grants. Eligible for assistance are manufacturing and extractive firms; starting from 2001, the L488 scheme has been extended through separate auctions to the tourism and transport sectors. Investment qualified for the intervention by the L488 are: setting-up, extension, modernization, restructuring, reactivation and relocation.

Three main features of L488 are important for the evaluation analysis:

- the L488 makes clear the targets of the policy intervention;
- the selection mechanism of L488 identifies projects that are viable but cannot be subsidized due to funds shortage;
- L488 operates at a regional level.

First of all, L488 is basically a national tender for incentives where the automatic allocation is based on general criteria expressing the policy preference. For a detailed description of Law L488 see, among others, Bronzini and de Blasio (2006), Carlucci and Pellegrini (2003) and Carlucci and Pellegrini (2005).

Incentives are allocated on the basis of regional competitive auctions. In each auction the investment projects are ranked on the basis of these five pre-determined criteria:

1. quota of owner capital invested in the project;
2. number of new employees per unit of investment;
3. ratio between the subsidy requested by the firm and the higher subsidy applicable, given the rules determined by area by the EU Commission;
4. a score related to the priorities of the region in relation to location, project type and sector;
5. a score related to the environmental impact of the project.

The criteria 4 and 5 were introduced starting from the 1998 (3rd auction). The five criteria carry equal weight: the values related to each criteria are normalized, standardized and added up to produce a single score that determines the position of the project in the regional ranking. The rankings are drawn up through the decreasing order of the score awarded to each project and the subsidies are allocated to projects until funding granted to each region is exhausted. There are also special rankings for large projects and reserved lists for small and medium-sized firms.

The five indicators are a clear expression of the policymakers' preferences. The share of the own funds invested in the project can be considered an (imperfect) proxy of the entrepreneur assessment of the project viability and success: higher the share, greater the commitment of the owner to the project (Chiri and Pellegrini (1995)). Moreover, the share is highly correlated to the economic and financial situation of the firm: the more profitable firms choose to assign a higher share of own funds to the project (Parascandolo and Pellegrini (2001)). Hence, the subsidized firms tend to be more profitable (and more efficient) than the non subsidized ones. The number of new job per unit of total investment is a central indicator, used to re-equilibrate the negative substitution effect of the capital subsidy to the firm labor demand. The policy makers express a preference for new projects and for labor-intensive investments. In order to increase the probability to receive the subsidy, the firms can choose to overshoot the optimal (i.e. the efficient) number of people to employ in the project.

5.3.1 The higher applicable subsidy

The amount of aid requested by firms, relative to ceilings established by the European Union, is the key indicator that transforms the allocation procedure to an auction

mechanism. The maximum ratio of subsidies on the total investment that a firm may request, depends on geographic areas and sizes. It can be easily summarized by Table 5.1, where S-M stands for small and medium size firms, while L stands for large firms. The quantity ESL stands for Equivalente Sovvenzione Lorda: it represents the equiva-

Typology	Subsidized areas	Subjects	Maximum grant	
			ESN	ESL
Objective 1: admitted areas to dispensation art. 87.3a UE Agreement	CALABRIA	S-M	50%	15%
		L	50%	/
	CAMPANIA PUGLIA BASILICATA SICILIA SARDEGNA	S-M	35%	15%
		L	35%	/
Admitted areas to dispensation art. 87.3c UE Agreement	MOLISE	S-M	20%	10%
		L	20%	/
Not admitted areas to dispensation art. 87.3c UE Agreement	MOLISE	S		15%
		M		7,5%
		L		/
Admitted areas to dispensation art. 87.3c UE Agreement	ABRUZZO	S-M	20%	10%
		L	20%	/
Objective 2	ABRUZZO	S-M		15%
		M		7,5%
		L		/

Table 5.1. Intensity of the subsidies

lent gross subsidy, that is the nominal subside expressed as the ratio between the total financed amount and the admitted investment value a part from the tax regime. ESN stands for Equivalente Sovvenzione Netta, that is the net benefit of enterprises: with respect to the ESL, it also consider the tax regime imposed to the firms.

The indicator aims to “reveal” the minimum amount of subsidy regarded by the firm as indispensable for the project realization. Therefore the firm can influence the likelihood of obtaining the incentive, self-reducing the “rent” granted by the subsidy, and the policy makers maximize the number of subsidized investments given the financial resources available, reducing the welfare losses due to a unique subsidy rate.

It is easy to understand that these thresholds affect the selection process, in particular the part of the selection process relative to the received level of treatment. To

be more clear consider a firm with a small dimension in the Campania: it never can request more than 35% of the total investment. It means that, in the matching procedure, it cannot be paired with a control unit belonging, for example, to another area. If an exact matching algorithm is chosen, this would be impossible. But if a propensity score approach is implemented, with a propensity function containing all the variables that affect the two selection processes, it might happen, also if the balancing property is satisfied. Then, we are in a case where there are some imposed restrictions on the treatment level assignment rule. As mentioned in the previous chapter, it might be the case of splitting the selection process in order to separately consider the participation decision and the treatment level assignment. In the following chapter these two approaches will be compared: a “full” propensity score function versus a selection process function composed of two related models. In other words, a constrained treatment level versus an unconstrained treatment level framework.

5.3.2 Why not a RDD approach?

Another interesting features of the auction mechanism is the presence of a set of firms willing to invest and having a valid investment project, as checked by an preliminary screening carried over by a set of appointed banks. They are admitted into the ranking, but they did not receive any subsidies because their scores were too low in the ranking. Consequently, for every auction, an own regional ranking and threshold is carried out. The presence of thresholds might invite to use a regression discontinuity design approach: near the threshold an experimental case might be reproduced and treated and non-treated units might be directly compared. But the presence of multiple ranks, depending on auctions and regions, implies unknown predetermined thresholds and reduces the applicability of the method because of the low number of observations near each cut-off point. However, this characteristic of L488 implies that firms, with the same propensity to be financed, can be subsidies or rejected from the auction on the base of their regional ranking.

The mentioned aspects reveal themselves an important features supporting policy evaluation analysis. The procedure uses the indicators as selection variables: therefore the indicators can explain most part of the differences between the group of subsidized and the group of non subsidized firms. This is of paramount importance for the construction of the counterfactual scenario in the evaluation analysis. Being the indicators observable, we can reconstruct the selection process, estimating the selection effect in

the control group. Moreover, the non subsidized firms are eligible to be part of a control group, as they show a propensity for investment and a need to invest which is very similar to that of subsidized firms (Brown M. A. and Elliott (1995), Carlucci and Pellegrini (2003) and Bronzini and de Blasio (2006)). The Law regulation also imposes that financing under L488 can not be combined with other source of public financing. Firms applying for the subsidies by L488 have to give up to other public subsidies, reducing the possibility of a double subsidization and the use of other public subsidies. The regional ranking produced by L488 reveals itself another important feature in implementing the evaluation analysis. Firms with the same level in the selection variables can be financed or not depending on the regional threshold. As a consequence, different groups of financed and not financed firms with respect to different regional thresholds are provided and, at the national level, an overlapping area of firms with the same propensity to be subsidized or not is available. This makes possible to compare firms with the same characteristics but with different selection results, as the matching evaluation technique requires. Therefore L488 mechanism allows to reproduced the treatment group among the non treated, re-establishing the experimental conditions in a non-experimental setting, and to construct the correct counterfactual for the evaluation analysis by matching method.

5.4 The data implementation

The data used in the analysis are collected in two different surveys: the L. 488 administrative dataset and the AIDA, which contains the budgets delivered by a subset of Italian firms to Chambers of Commerce. The integration among the different sets of data has required a complex process of cleaning and merging.

First of all, the identification of the eligible projects is carried out. The financed projects group (treated group) consists of all the “winning” (i.e. funded) projects in Mezzogiorno (ob. 1 regions, excluding Abruzzo) according to the ranks of all regional auctions. Projects are eligible for control group if are in the manufacturing sectors and if are admitted to evaluation to the regional auctions but non financed (“losers”). Projects that were funded in other auctions (special actions dedicated to Northern regions, or to areas devastated by an earthquake or to tourism and retail sectors) or via special regional ranks were discarded from control group. Projects applied for more than one L. 488 auction and financed once time have also been excluded from the control group.

Particular attention was dedicated to the selection of the projects that did not

present anomalies and irregularities for the analysis: anomalous projects were intentionally discarded, in order to not affect the results by programs inherited from previous instruments or by projects still in progress. Financed projects, whose investment program have not yet concluded, have been discarded. Both treated and control group were subsequently cleaned out of the projects which year of conclusion (actual and scheduled, respectively) has preceded the year of the auction.¹ Another group of discarded projects is represented by the programs started (or scheduled to start) before the year preceding the publication of the auction. Since their activation cannot be directly linked to L. 488, these projects must be regarded as anomalous. Finally, all the projects started (or scheduled to start) after 1999 have been discarded: the choice has been motivated by the impossibility to evaluate the projects, being missed project information after a sufficient temporal lag following its conclusion.

The third step regards the integration of L. 488 dataset with AIDA budgetary data. A more detailed description of this dataset integration can be found in Appendix A.

After verifying that the cleaning and integration procedures do not have a different impact on financed projects and control group, the attention focused on the final dataset on which the evaluation model has been implemented. It consists of 665 financed projects and 1.228 non financed projects for the years 1996-2000.

For the validation of the control group, a comparison of the main characteristics of the projects and of the ranking indicators in the sample of subsidized and non subsidized firms are presented. Similarities in the two sample referring to budgetary data, both for year before the start of the project (year 0) and for the year following the conclusion (year 1) of the investment program are analyzed. The results are presented in tables 5.2 and 5.3².

As far as the distribution of projects is concerned, a substantial homogeneity between the two groups is found, according to region, firm dimension, economic activity sector and investment typology. Some difference are found for what concerns large firms, which represents 11,2% in the control group and 18,5% in the treatment group; similarly the share of small enterprises turns out to be smaller in the financed projects group (54,1% vs. 66,8%). Regarding economic activity sector, the most difference regards food industry which share is double in the control group (8,6% vs. 19,4%). The distribution according to project typology does not introduce remarkable differences.

¹The incongruity derives from the fact that in the first two auctions (years 1996 and 1997) the L. 488 has inherited applications from the previous incentive instrument (law 64/1986), that was closed in 1992.

²Source: elaboration of Pellegrini and Centra (2006) on L488 and AIDA data. We have deflated

The levels of the indicators, crucial parameters for the application of the model, show a substantial homogeneity between the two groups. The analysis of the indicators median values does not indicate strong differences between the two sample in the budgetary data. The subsidized firms are slightly greater, more profitable and more capital intensive, as expected, than the non subsidized firms.

As regards the different time span of the auctions, the dataset contains information on the years of the effective beginning and the end of the investment project. Exploiting this information, we can estimate the impact of the subsidy confronting the balance sheet the year before the project is really started and the year after the investment is really concluded³. The information is relevant, because the L. 488 procedure requires neither the investment project to be actually started by the time of the first subsidy instalment, either to be over after two years since the beginning (in this case, however, also the payments of the following instalments are lagged). In the analysis, the time span can differ for each project, depending on the beginning and the end of the investment. A correct matching procedure requires to impute an ending date also for the not subsidized firms. The adopted hypothesis is that the ending date is equal to the date scheduled for the investment starting augmented by the average investment period by auction calculated for the subsidized firms sample.

profitability and leverage variables by the investment deflator (base=2000)

³This is the date corresponding to the end of the inspection carried out by the appointed bank. Therefore, it can overestimate the length of the investment period by 6-12 months.

		Projects			Distribution %		
		Not financed	Financed	Total	Not financed	Financed	Total
Geographical location	Basilicata	82	28	110	6,68%	4,20%	5,81%
	Calabria	107	41	148	8,71%	6,16%	7,82%
	Campania	418	232	650	34,04%	34,83%	34,34%
	Molise	31	49	80	2,52%	7,36%	4,23%
	Puglia	294	130	424	23,94%	19,52%	22,40%
	Sardegna	87	38	125	7,08%	5,71%	6,60%
	Sicilia	209	147	356	17,02%	22,07%	18,81%
	Total	1228	666	1893	100,00%	100,00%	100,00%
Firm dimension	Large	138	123	261	11,24%	18,50%	13,79%
	Medium	270	182	452	21,99%	27,37%	23,88%
	Small	820	360	1180	66,78%	54,14%	62,33%
	Total	1228	665	1893	100,00%	100,00%	100,00%
Economic activity sector	Mining and quarrying of energy producing materials	0	1	1	0,00%	0,15%	0,05%
	Mining and quarrying, except of energy producing	31	10	41	2,52%	1,50%	2,17%
	Manufacture of food products, beverages and tobacco	238	57	295	19,38%	8,57%	15,58%
	Manufacture of textiles and textile products	62	56	118	5,05%	8,42%	6,23%
	Manufacture of leather and leather products	34	33	67	2,77%	4,96%	3,54%
	Manufacture of wood and wood products	24	16	40	1,95%	2,41%	2,11%
	Manufacture of pulp, paper and paper products	75	46	121	6,11%	6,92%	6,39%
	Manufacture of coke, refined petroleum products	43	24	67	3,50%	3,61%	3,54%
	Manufacture of chemicals, chemical products and	55	41	96	4,48%	6,17%	5,07%
	Manufacture of rubber and plastic products	74	62	136	6,03%	9,32%	7,18%
	Manufacture of other non-metallic mineral products	138	68	206	11,24%	10,23%	10,88%
	Manufacture of basic metals and fabricated metal	166	88	254	13,52%	13,23%	13,42%
	Manufacture of machinery and equipment n.e.c.	62	33	95	5,05%	4,96%	5,02%
	Manufacture of electrical and optical equipment	55	36	91	4,48%	5,41%	4,81%
	Manufacture of transport equipment	56	31	87	4,56%	4,66%	4,60%
	Manufacturing n.e.c.	69	37	106	5,62%	5,56%	5,60%
	Unclassified service activities	46	26	72	3,75%	3,91%	3,80%
	Total	1228	665	1893	100,00%	100,00%	100,00%
Project typology	Modernization	217	92	309	17,67%	13,83%	16,32%
	Extension	518	353	871	42,18%	53,08%	46,01%
	Setting up	421	188	609	34,28%	28,27%	32,17%
	Reactivation	1	3	4	0,08%	0,45%	0,21%
	Reconversion	3	0	3	0,24%	0,00%	0,16%
	Restructuring	67	29	96	5,46%	4,36%	5,07%
	Relocation	1	0	1	0,08%	0,00%	0,05%
	Total	1228	665	1893	100,00%	100,00%	100,00%

Table 5.2. Distribution of projects according to main characteristics (1996-2000)

		Median		
		Not financed	Financed	Total
Investment(Euro)*		1358452	1415674	1372096
Ranking Indicator	Own capital indicator	0,448	0,560	0,484
	Employment indicator	0,003	0,004	0,003
	Subsidy share indicator	1,333	1,351	1,333
Year 0	Turnover	3.541.644	4.125.601	3.746.266
	Employees	25	30	26
	Fixed assets	1.244.287	1.537.239	1.302.372
	Gr. margin / Turnover	0,09	0,10	0,09
	ROI	4,20	5,57	4,58
	ROE	4,02	6,99	5,05
	Fin. charges / Turnover	0,03	0,03	0,03
	Turnover / Employees	169.380	159.485	165.399
	Fin. charges / Debt	0,05	0,05	0,05
Year 1	Value Added	928.895	1.203.206	1.000.404
	Turnover	4.680.689	5.863.183	5.090.816
	Employees	29	42	32
	Fixed assets	1.984.960	3.054.411	2.412.743
	Gr. margin / Turnover	0,09	0,10	0,09
	ROI	3,47	3,26	3,37
	ROE	2,61	1,93	2,30
	Fin. Charges / Turnover	0,02	0,03	0,02
	Turnover / Employees	185.048	154.387	169.901
	Fin. Charges / Debt	0,03	0,03	0,03
	Value Added	1.135.065	1.778.691	1.335.058

Table 5.3. Summary of main covariates in the final dataset (1996-2000)

Chapter 6

Application

6.1 Introduction

This part of the thesis show the results of the empirical application: the impact of subsidies allocated by law 488/92 in Italy on subsidized firm are estimated using the proposed nonparametric approach. The continuous treatment setting is appropriate in the case of firm incentive programs, where firms receive different amounts of subsidies. Even more, in this situation, it is not so unlikely to find non random treatment level assignment. That is, we are away from an experimental data framework because there is a non random selection process not only on the participation decision but also with respect to the treatment level assignment. This means law 488 determines a deliberate selection process: the subsequent selection bias problem have been tackled using the proposed estimation method. The empirical findings are the core of this chapter.

6.2 The treatment variable: amount of subsidies

As we said time after time, the treatment variable plays a fundamental role in the evaluation program in a continuous treatment contest. Then, the concern of this section is on the continuous treatment variable of the empirical application. As mentioned before, the law 488 is a form of private capital accumulation: it consists of subsidies to capital accumulation. Then, the continuous treatment variable of the study might be the amount of these subsidies, received by treated firms. However, a simple descriptive analysis of this variable (Table 6.1), may reveal some limitation by adopting this as

treatment variable. By construction, it depends on the investment the subsidies refer to: this implies an heterogeneous distribution of its values, depending mainly on the size of the firms.

Granted Subsidies (Euros)				
	Percentiles	Smallest		
1%	44340.23	20965.89		
5%	103767.5	24106.71		
10%	167532.8	35717.29	Obs.	665
25%	309356.7	38134.29	Sum of wgt.	665
50%	587821.9		Mean	1440201
		Largest	Std. Dev.	5896441
75%	1187916	1.46e+07		
90%	2599559	1.47e+07	Variance	3.48e+13
95%	4428788	1.62e+07	Skewness	21.27131
99%	1.37e+07	1.44e+08	Kurtosis	509.3261

Table 6.1. Summary of the granted subsidies

In order to reduce this source of variability and to limit the range of the treatment variable, what is proposed here is to use the ratio of the subsidies on the investment. In this way the treatment variable takes value in the interval $[0,1]$. Throughout the analysis this indicator will be called *quota*. Table 6.2 and Figure 6.1 report a descriptive and a graphical analysis of this variable.

Percent share of subsidies on the investment				
	Percentiles	Smallest		
1%	14.018	8.595		
5%	21.932	9.825		
10%	27.289	9.827	Obs.	665
25%	37.538	11.462	Sum of wgt.	665
50%	45.921		Mean	45.478
		Largest	Std. Dev.	13.132
75%	54.013	77.308		
90%	61.941	78.241	Variance	172.448
95%	66.690	78.513	Skewness	-0.086
99%	75.731	94.637	Kurtosis	3.164

Table 6.2. Summary of the percent share of subsidies on the total investment

The distributional graph of the treatment variable shows that the share of requested

subsidies on the total investment seems to have a symmetric distribution over its mean value equal to 46%.

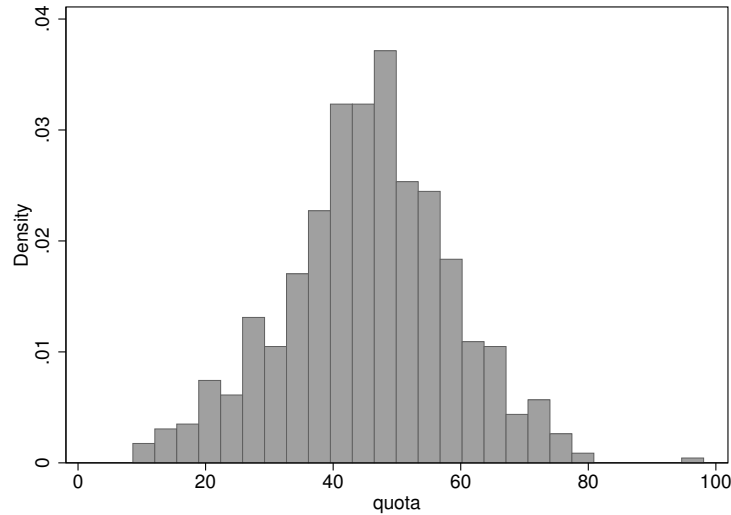


Figure 6.1. Distribution of the treatment variable

6.3 The outcome variables

As mentioned before, the aim of this application is to provide a statistically robust evaluation of the impact of subsidies to capital accumulation on subsidized firms. In particular, we want to evaluate if the receipt of financial assistance from public funds actually makes a difference in the firm performances in terms of investment, new employment, profit and labor productivity. Then, as outcome variables, on which we will estimate the treatment effects, four set of variables are considered: firm growth (turnover, number of employees, fixed assets), profitability (gross margin/turnover), productivity (per capita turnover) and leverage (debt charges/turnover). For turnover, employment, fixed assets and per capita turnover estimates of treatment effects are computed as the difference in (weighted average) growth rates between subsidized and non subsidized firms, for the other variables as the difference in levels. Results are based on the not subsidized firms sample where the ending year for the non subsidized firms is imputed on the base of the average investment length by auction in the subsidized firms sample. The presence of several anomalous data (as signalled by the large difference

	Obs.		Mean	
	Financed	Not Financed	Financed	Not Financed
Turnover	493	780	37.95	34.32
Employment	517	742	49.94	42.51
Fixed Assets	521	776	116.40	97.35
Gr. Margin/Turnover	422	675	0.000036	0.000582
Per Capita Turnover	501	753	5.00	11.38
Debt Charges/Turnover	503	759	-0.00559	-0.00533

Table 6.3. Summary of outcome variables

between median and mean across indicators), suggests to select only firms with non-negative values for turnover, employment and assets, and to trim the subsidized and non subsidized firm samples at the 5 and 95 percentiles¹. Table 6.3 and Figures from B.1 to B.6 in Appendix B report, for this selected sample, mean values of these outcome variables, by treatment status. The selected sample of firms has a different distribution for the outcome variables with respect to the treatment status. In particular, financed firms seems to have higher values in terms of firm growth, substantially equal values for profitability and leverage, and lower values for productivity. By applying our methods, we will provide a statistical robust evaluation of this heterogeneity between the two groups: we will be able to measure how much part of this difference is caused by the program participation.

6.4 Structured from of the selection process

The first step in the evaluation procedure is the specification of the propensity score model identifying the participation decision process. We adopted a logit specification of the treatment variables. For the covariates identification, we take advantage of the selection mechanism that is used to allocate the incentives under the L. 488 by including the selection indicators in the equation of the propensity score. The main ranking indicators are introduced, either in level or squared and cubed. Dummy variables relative to the different auctions are considered because they reflect some specific characteristics especially relative to the admission of projects already concluded. Moreover, the interaction between main indicators and dimension (large dimension by European Union

¹A similar procedure has been applied by Bronzini and de Blasio (2006)

Commission definition) is introduced in the model specification. The regional and sectorial indicators are also considered. The first set of indicators allows to control for differences in regional ranking and threshold; the sectorial dummies take into account of productive heterogeneity of firms attending to auctions. The final specification of the logit model for propensity score and the parameter estimates are in Table 6.4. To

LR chi2	1065.59							
Prob	0.000							
Pseudo R2	0.4349							
Obs	1888							

variables	Coef.	Std. Err.	variables	Coef.	Std. Err.	variables	Coef.	Std. Err.
ind1	20.14	5.87	large*ind1	-5.16	1.41	ateco DJ	0.31	0.23
ind2	225.96	31.53	large*ind2	31.84	45.14	ateco DK	0.40	0.32
ind3	3.64	0.39	area1	1.73	0.50	ateco DL	0.56	0.33
ind1_2	-33.28	15.14	area2	1.54	0.43	ateco DM	0.08	0.37
ind1_3	28.64	12.35	ateco DB	0.24	0.29	ateco CB	-0.24	0.46
ind2_2	-2844.27	885.92	ateco DC	0.03	0.38	moderniz	1.22	1.05
large	2.56	0.79	ateco DE	0.31	0.34	extension	1.70	1.03
auction 1	5.93	0.58	ateco DF	0.62	0.37	new	1.85	1.03
auction 2	3.27	0.23	ateco DG	0.84	0.34	restructur	1.59	1.09
auction 3	-0.98	0.17	ateco DH	0.96	0.29	cons	-16.21	1.62
med*ind1	0.70	0.33	ateco DI	0.65	0.24			

Table 6.4. Structured form: Propensity score estimate

test the "balancing hypothesis" we have followed the procedure proposed in Ichino and Becker (2002). Splitting the sample by propensity score in 7 blocks, we tested that the balancing hypothesis is satisfied.

The second step of the evaluation procedure is the specification of the treatment level model. As mentioned before, to consider for the relation between participation decision and treatment level assignment, different specification for the share of subsidies are estimated at different point of the propensity score distribution. In particular, we split the sample by equal average propensity score values between treated and non treated units and estimate a linear regression model for the share of subsidies, for each chosen blocks of the propensity score. In other words, we use the 7 blocks identified by the balancing hypothesis procedure for the propensity score estimation. In order to select which covariates to include in the analysis the standard goodness of fit statistics together with the common specification model tests have been carried out. Some covariates are common for each model: they mainly refer to the variable reflecting the subsidies

limitation imposed by the law (firm size and areas). While other variables may be found in some specification but not in others: they are the sectorial dummies, the kind of investment qualified for the intervention or the amount of debts charges on debts stock. Table 6.5 reports the linear regression estimates for the 2-nd block of the propensity score. Given the estimated coefficients of the models and the values of covariates, the

Number of obs	42		
Prob > F	0.000		
Adj R-squared	0.732		
variables	coef	Std.Err	P> t
medium	0.106	0.029	0.001
auction 3	-0.071	0.026	0.010
large	-0.141	0.043	0.003
ateco DN	-0.270	0.075	0.001
ateco DB	0.197	0.055	0.001
ateco DJ	0.094	0.031	0.004
new	0.147	0.035	0.000
auction 2	0.149	0.036	0.000
extension	0.102	0.029	0.001
area 1	0.073	0.039	0.067
cons	0.422	0.035	0.000

Table 6.5. Structured form: Level of treatment estimate

predicted values were computed also for the group of the non-participants, in order to use this variable to compute the second step of the proposed matching procedure. To test the "balancing hypothesis" we have followed the same procedure proposed in Ichino and Becker (2002) properly adapted to the specific case. We tested that the balancing hypothesis is satisfied for each model.

6.4.1 Impact of the Law 488

Once estimated the two models we computed the two-step matching (diff-in-diff) procedure. Among the matching with replacement methods proposed in literature we have chosen, for each step, the radius matching, with four different size for the radius:

- a predetermined radius equal to the 10% of the propensity score (or function) range (that is equal to 0,1)
- radius equal to the mean of all distances between treated and non treated units

radius=10%*					
Outcomes	Subs.	Not Subs.	TTE	Std. Er.	T-test
Turnover	410	592	16.811	9.412	1.786
Employment	430	555	24.796	11.017	2.251
Fixed asset	442	591	39.898	26.318	1.516
Gr. Margin/Turnover	342	503	-0.009	0.008	-1.118
Per capita Turnover	416	568	-6.373	7.840	-0.813
Debt Charges/Debt Stock	435	573	0.004	0.003	1.567

*radius equal to 10% of the propensity score range

Table 6.6. Structured form: TTE estimates radius=fix

radius=mean*					
Outcomes	Subs.	Not Subs.	TTE	Std. Er.	T-test
Turnover	490	592	8.893	3.728	2.385
Employment	514	555	10.995	4.503	2.442
Fixed asset	518	591	27.470	10.128	2.712
Gr. Margin/Turnover	417	503	0.0002	0.004	0.047
Per capita Turnover	500	568	-5.312	3.091	-1.718
Debt Charges/Debt Stock	501	573	0.0005	0.001	0.385

*radius equal to the mean distance

Table 6.7. Structured form: TTE estimates radius=mean

- radius equal to the maximum of the minimum of all distances
- radius equal to the standard deviation of all distances between treated and non treated units

Then average treatment level effects (ATLE) are estimated by equation (4.6). In order to have a general overview of these effects, Tables from 6.6 to 6.9 reports the “total” average treatment effect estimates for different kind of radius sizes. They have the same interpretation as average treatment effects (TTE) in the traditional binary case.

As expected, the growth impact of 488 on subsidized firms is positive and statistically significant: the turnover increases from 9 to 12 points higher in the subsidized firms than in non subsidized ones, depending on the radius of the matching algorithm; the number of employees is from 11 to 25 per cent points higher and fixed assets increases to 27 per cent points. These results are all statistical significant. Being the average time span

radius=maxmin*					
Outcomes	Subs.	Not Subs.	TTE	Std. Er.	T-test
Turnover	438	592	20.185	11.978	1.685
Employment	465	555	27.663	14.198	1.948
Fixed asset	473	591	39.321	33.747	1.165
Gr. Margin/Turnover	374	503	-0.011	0.011	-0.980
Per capita Turnover	460	568	-8.759	10.286	-0.852
Debt Charges/Debt Stock	451	573	0.005	0.003	1.393

*radius equal to the maximum of the minimum of all distances

Table 6.8. Structured form: TTE estimates radius=std

radius=std*					
Outcomes	Subs.	Not Subs.	TTE	Std. Er.	T-test
Turnover	471	592	12.195	5.080	2.401
Employment	497	555	16.496	6.283	2.625
Fixed asset	501	591	26.065	14.401	1.810
Gr. Margin/Turnover	400	503	-0.002	0.005	-0.359
Per capita Turnover	480	568	-5.615	4.267	-1.316
Debt Charges/Debt Stock	183	573	0.001	0.002	0.916

*radius equal to the standard deviations of all distances

Table 6.9. Structured form: TTE estimates radius=mean

equal to around 3,5 years, the additional annual employment growth rate imputed to L. 488 fluctuates from 3,1 to 7,1 percent points. The impact on growth in fixed assets is also very high (an annual average of more than 7,7 points). In general, the results confirm the findings reported in Carlucci and Pellegrini (2003) using a parametric approach and the ones of Bernini *et al.* (2006) using a matching approach but in a binary case. Moreover, they does not necessarily contradict the presence of investment intertemporal substitution shown by Bronzini and de Blasio (2006), even if the amount of additional capital installed in subsidized firms in the period is large. However, for fixed assets the statistical significance of the incentive additional effect is lower.

The effects of the 488 are in line whit the (explicit or less explicit) targets: the subsidized firms have invested more than the non subsidized ones, achieving more turnover, employment and fixed assets. The question is if the subsidized firms have also increased

their efficiency, measured by productivity and profitability, in order to maintain a positive long run growth rate. Profitability is measured by gross margin on turnover. The results show a slight negative difference for the subsidized firms, but the difference is not statistically significant. However, the total amount of profits, as turnover, is increased faster in subsidized firms. Productivity is proxied by per capita turnover. The impact of L. 488 is negative but not statistically significant. This negative average effect turns out also in the work of Bernini *et al.* (2006) where a matching approach is followed. However here the result is statistically significant: the labor productivity growth is 4-8% higher in non subsidized firms. The authors give some explanation for this productivity gap. They argue that if the investment productivity curve is decreasing, the reduction in the investment cost generated by the subsidy drives the subsidized firm to invest in projects with a lower than average productivity. Furthermore, if the option for a given sector is investing or restructuring, the not subsidized firms can have chosen to restructure, increasing the productivity, whereas the subsidized firm can have chosen to invest, increasing production and employment. The fact that our method turns out a not significant result can be attributed to the two step matching procedure: when only one matching is computed, that is in the traditional binary treatment case, it may be the case that pairs are computed among units which are more distant in term of their propensity score values than in the double matching case. This can reduce the significance of the average treatment effect in this latter case.

As expected, the debt cost on turnover is slightly higher in the subsidized than in the not subsidized firms, but the average effects is not statistically significant. The reason can be imputed to an increase in debt that the subsidized firms have had to face in order to finance the new investment, whereas the non subsidized firms could have refrained from investing. However, the subsidy has not radically changed the financial state of the firm.

These results show only an average tendency of treatment effects, but it is easy to understand that the effects of L. 488 can be non homogeneous by firm. On the basis of the estimated ATLE, our method allows to compute the impact of the amount of subsidy on the treatment effect. In order to investigate if the treatment level differently affects the response variable, we use an OLS estimator imposing a quadric relation between effects and level of subsidies. We restrict the analysis only for outcome variables with a significant average treatment effect (see Tables from 6.6 to 6.9). Estimates are reported in Table 6.10. The results show a significant quadratic relation between effects and treatment levels for turnover and employment. In particular, given the sign of the coefficients for the level variable (quota), the analysis evidences an increasing positive

	Employment			Turnover		Fixed Assets
	radius fix	radius mean	radius std	radius mean	radius std	radius mean
quota	2.998*	3.439*	3.769*	1.159	3.348*	0.108
quota ²	-0.035*	-0.034*	-0.039*	-0.012	-0.040*	-0.009
const	-31.91	-68.55*	-65.86*	-15.38	-49.39*	40.18
Obs.	58	65	61	65	58	64
Ad. R^2	0.089	0.201	0.127	-0.003	0.187	0.008
Prob. F	0.029	0.000	0.007	0.405	0.001	0.290

*p<0.05

Table 6.10. OLS estimates: impact of the amount of subsidies on the treatment level (structured case)

impact of the amount of incentives for low values of treatment level with respect to non treated firms. After reaching a maximum level, the effect of the subsidies on the outcomes is decreasing. This tendency can be better viewed by graphing the predicted values of these OLS estimates (Figure 6.2). Apart for fixed assets, that returns not significant coefficients, the analysis evidences an increasing positive impact of the amount of incentives until the grant is equal to 40-50% of the total investment. After this peak, the effect of the subsidies on the outcomes is decreasing: firms investing in the project less than the half of their capital achieve less favorable performance. Then the results evidence that the additional effect of large grants with respect to the level of the firm's investment is low.

In order to better detect some heterogeneity of the effects with respect to the level of treatment, that is consistently with the recent literature on the heterogeneity of the effects in the program evaluation contest (see among others Athey and Imbens (2006)), we also estimate this relation by a quantile regression. Estimates of the quantile regression are reported in Tables 6.11 and 6.12.

The ability of quantile regression models to characterize the heterogeneity impact of variables on different points of an outcome distribution makes them appealing for evaluating the effects of policy interventions. Restating our problem of evaluating the effects of policy interventions in a quantile regression framework allows us to investigate if treatment groups have been different benefited from the treatment and to provide an analytical description of the effects distribution with respect to the amount of granted subsidies. The results confirm a significant quadratic relation between average effects

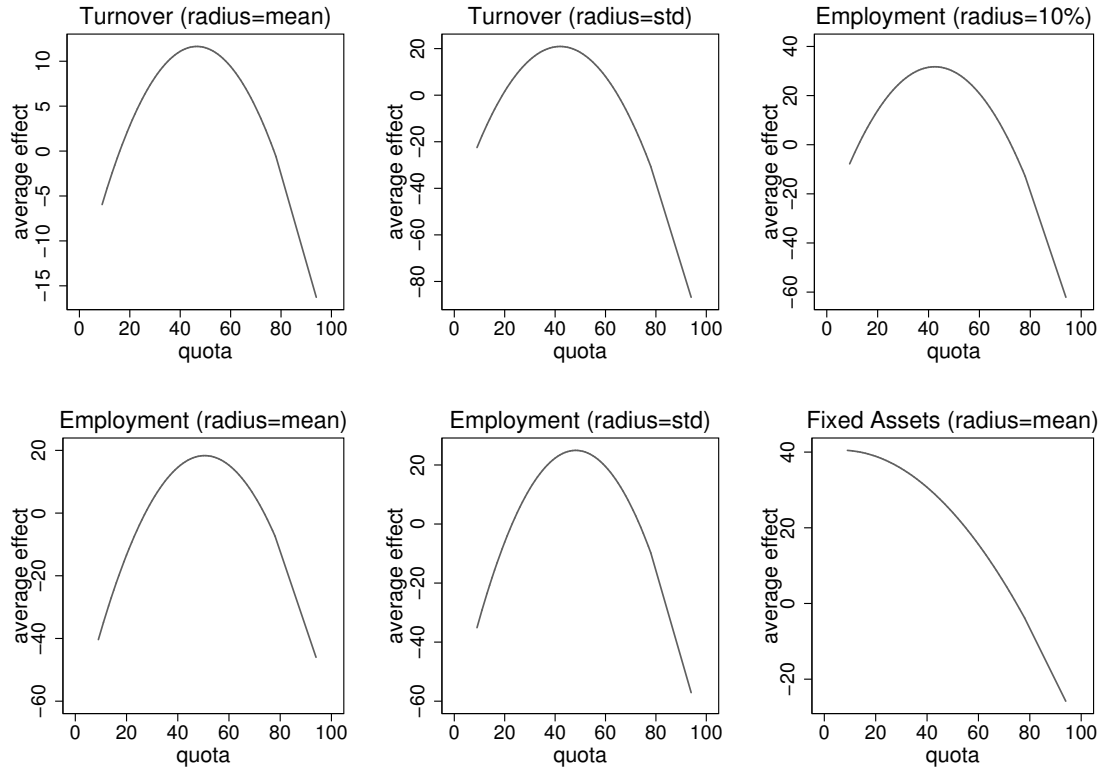


Figure 6.2. Predicted OLS estimates for treatment impact on the amount of subsidy (structured case)

and levels for low percentile values of the dependent variable, especially for employment.

This tendency is confirmed also by adopting a non parametric estimator for the relation between effects and treatment levels. Figures 6.3 and 6.4 show the results of the non parametric mean regression estimator (using the Nadayara-Watson kernel estimator, Nadayara (1964)) for the outcome variables.

Again, the results confirm heterogeneity of the treatment outcome with respect to different levels of treatment, especially for the outcome variable relative to employment growth. In particular, it is detected an increasing relation for low values of subsidies; after a peak is reached, it follows a decreasing relation. This profile is robust to the radius size adopted in the matching procedure. For turnover growth, this kind of relation is detected only choosing a radius equal to the mean distance. The value on which this maximum refers to is different for the two outcome variables: for turnover it is between

		Employment					
		radius fix	Pseudo R^2	radius mean	Pseudo R^2	radius std	Pseudo R^2
q5	quota	8.898*	0.500	6.051*	0.373	8.543*	0.536
	quota ²	-0.104*		-0.068*		-0.093*	
	const	-174.5*		-141.7*		-191.9*	
q25	quota	5.436*	0.252	5.990*	0.255	4.759	0.197
	quota ²	-0.060*		-0.063*		-0.052*	
	const	-104.9*		-133.9*		-96.19	
q50	quota	2.341	0.086	3.858*	0.119	2.377	0.031
	quota ²	-0.033		-0.038*		-0.028	
	const	-9.5		-81.7*		-31.76	
q75	quota	-3.773	0.018	1.268	0.069	-0.052	0.009
	quota ²	0.002		-0.008		0.002	
	const	47.81		-22.9		25.7	
q95	quota	-3.174	0.090	2.667	0.047	2.345	0.011
	quota ²	0.025		-0.023		-0.023	
	const	158.3		-17.3		2.6	

*p<0.05, bootstrapped standard errors

Table 6.11. Quantile estimates: Employment (structured case)

25 and 35% of the share of subsidies on the investment. For employment the impact is increasing until the grant is equal to 50-60% of the total investment. For fixed assets the relation between average effects and treatment level seems to be not significant, as confirmed by applying the parametric approaches above.

		Turnover				Fixed assets	
		radius mean	Pseudo R^2	radius std	Pseudo R^2	radius mean	Pseudo R^2
q5	quota	4.415*	0.288	8.008*	0.428	7.058*	0.108
	quota ²	-0.049*		-0.098*		-0.075*	
	const	-104.5*		-161.5*		-190.9*	
q25	quota	3.084*	0.159	3.063	0.177	3.041	0.049
	quota ²	-0.035*		-0.039		-0.036	
	const	-64.3*		-51.5		-65.7	
q50	quota	1.507	0.030	2.770	0.107	1.152	0.010
	quota ²	-0.016		-0.037		-0.014	
	const	-23.2		-33.1		1.7	
q75	quota	-1.377	0.015	-0.647	0.033	-3.879	0.030
	quota ²	0.017		0.003		0.038	
	const	41.5		47.71		143.5	
q95	quota	-2.497	0.147	-0.889	0.101	-12.500	0.145
	quota ²	0.033		0.007		0.108	
	const	83.5		68.5*		443.7	

*p<0.05, bootstrapped standard errors

Table 6.12. Quantile estimates: Turnover and fixed assets (structured case)

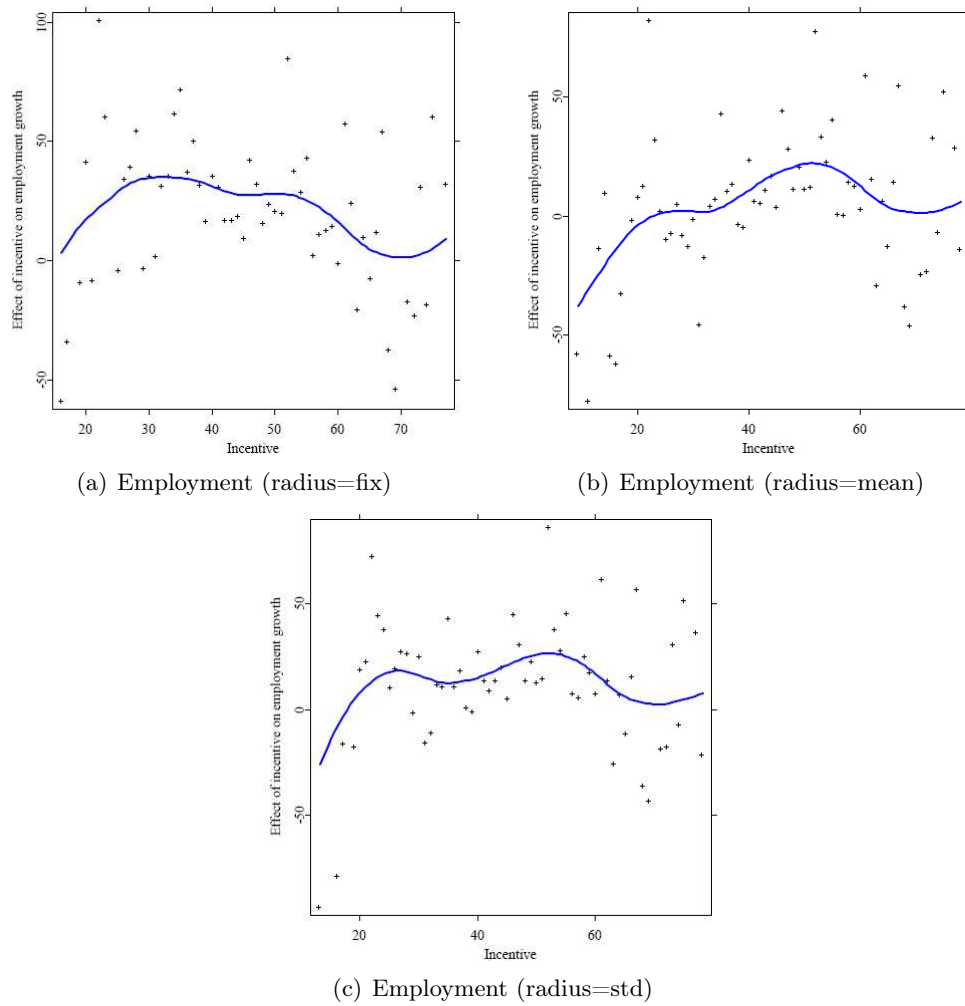


Figure 6.3. Kernel estimates for treatment impact on the amount of subsidy (structured case)

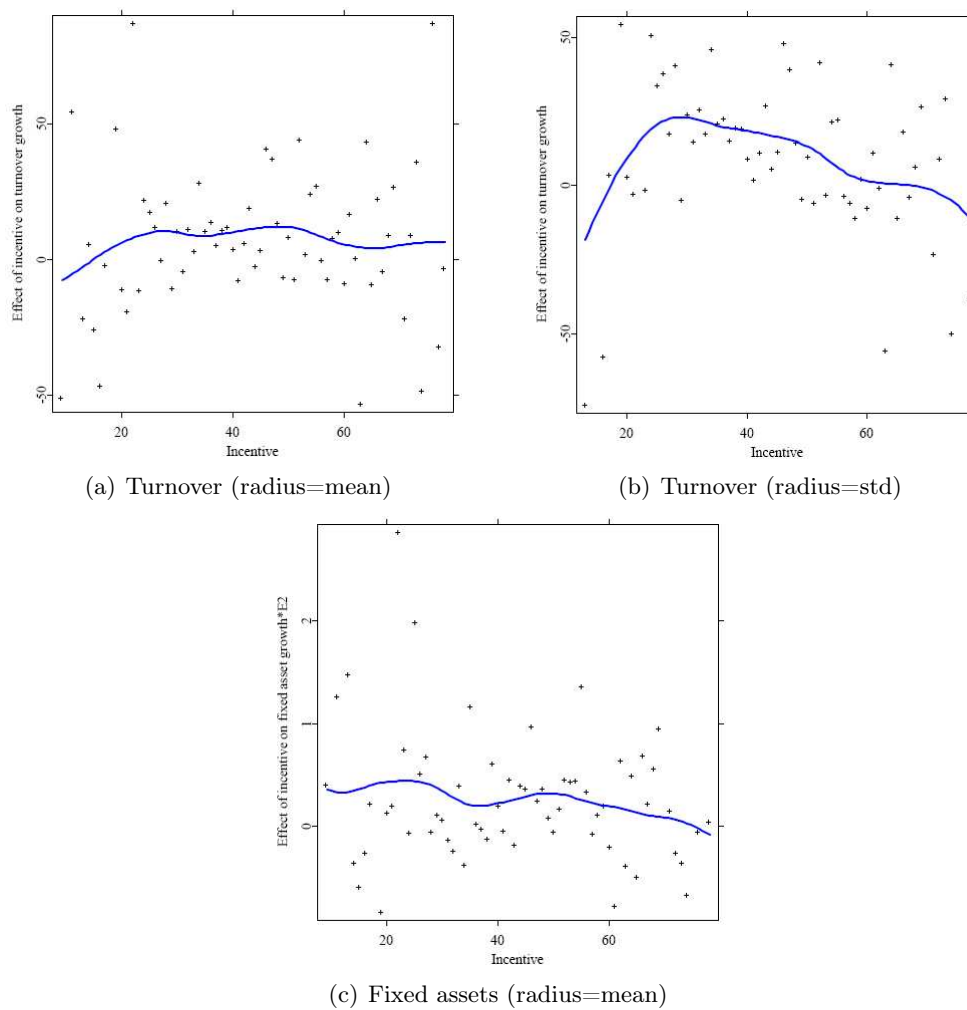


Figure 6.4. Kernel estimates for treatment impact on the amount of subsidy (structured case)

6.5 Non-structured from of the selection process

By adopting a unique specification for the selection process, we include in the model all the covariates entered in the two steps of the previous section (6.4). Again, for the propensity score model we adopted a logit specification of the binary treatment variables. The parameter estimates are in Table 6.13.

LR chi2	1005.04								
Prob	0.000								
Pseudo R2	0.4345								
Obs	1763								

variables	Coef.	Std. Err.	variables	Coef.	Std. Err.	variables	Coef.	Std. Err.
ind1	21.59	6.10	large*ind1	-5.545	1.44	ateco DJ	0.39	0.24
ind2	219.93	32.72	large*ind2	27.01	46.21	ateco DK	0.43	0.33
ind3	3.48	0.40	area1	1.84	0.51	ateco DL	0.62	0.35
ind1_2	-37.73	15.72	area2	1.59	0.44	ateco DM	0.08	0.39
ind1_3	32.65	12.80	ateco DB	0.31	0.30	ateco CB	-0.20	0.46
ind2_2	-2630.06	958.95	ateco DC	0.11	0.38	moderniz	1.61	1.23
large	2.82	0.81	ateco DE	0.32	0.34	extension	2.13	1.21
auction 1	5.94	0.60	ateco DF	0.69	0.37	new	2.31	1.21
auction 2	3.24	0.23	ateco DG	0.91	0.36	restructur	1.97	1.26
auction 3	-0.96	0.18	ateco DH	1.02	0.30	debt charges/debt	-0.07	0.55
med*ind1	0.57	0.33	ateco DI	0.74	0.25	cons	-16.60	1.77

Table 6.13. Non structured form: Propensity score estimate

To test the "balancing hypothesis" we have followed the procedure proposed in Ichino and Becker (2002). Splitting the sample by propensity score in 10 blocks, we tested that the balancing hypothesis is satisfied.

6.5.1 Impact of the Law 488

Once estimated the propensity score model we computed the matching (diff-in-diff) procedure. Again, among the matching with replacement methods proposed in literature we have chosen, for each step, the radius matching, with four different size for the radius, as specified above. Then average treatment level effects (ATLE) are estimated by equation (4.8). In order to have a general overview of these effects, Tables from 6.14 to 6.17 reports the "total" average treatment effect estimates for different kind of radius sizes. They have the same interpretation as average treatment effects (TTE) in the traditional binary case.

radius=10%*					
Outcomes	Subs.	Not Subs.	TTE	Std. Er.	T-test
Turnover	493	614	22.175	11.492	1.930
Employment	517	576	29.242	13.161	2.222
Fixed asset	521	610	37.923	29.964	1.266
Gr. Margin/Turnover	422	525	-0.009	0.010	-0.871
Per capita Turnover	501	594	-6.286	9.149	-0.687
Debt Charges/Debt Stock	503	596	0.005	0.003	1.430

*radius equal to 10% of the propensity score range

Table 6.14. Non-structured form: TTE estimates radius=fix

radius=mean*					
Outcomes	Subs.	Not Subs.	TTE	Std. Er.	T-test
Turnover	493	614	8.598	3.555	2.419
Employment	517	576	11.410	4.269	2.673
Fixed asset	521	610	24.180	9.368	2.581
Gr. Margin/Turnover	422	525	0.001	0.003	0.344
Per capita Turnover	501	594	-5.220	2.892	-1.805
Debt Charges/Debt Stock	503	596	0.001	0.001	0.771

*radius equal to the mean distance

Table 6.15. Non-structured form: TTE estimates radius=mean

radius=maxmin*					
Outcomes	Subs.	Not Subs.	TTE	Std. Er.	T-test
Turnover	493	614	23.869	14.707	1.623
Employment	517	576	32.484	17.057	1.904
Fixed asset	521	610	46.157	38.573	1.197
Gr. Margin/Turnover	422	525	-0.013	0.014	-0.959
Per capita Turnover	501	594	-8.333	11.914	-0.699
Debt Charges/Debt Stock	503	596	0.006	0.004	1.412

*radius equal to the maximum of the minimum of all distances

Table 6.16. Non-structured form: TTE estimates radius=std

Outcomes	radius=std*		TTE	Std. Er.	T-test
	Subs.	Not Subs.			
Turnover	493	614	11.291	4.767	2.396
Employment	517	576	15.894	5.801	2.732
Fixed asset	521	610	20.905	12.689	1.647
Gr. Margin/Turnover	422	525	-0.000	0.005	0.106
Per capita Turnover	501	594	-6.160	3.957	-1.557
Debt Charges/Debt Stock	503	596	0.001	0.001	0.843

*radius equal to the standard deviations of all distances

Table 6.17. Non-structured form: TTE estimates radius=mean

As above, the growth impact of 488 on subsidized firms is positive and statistically significant: the turnover increases from 9 to 11 points higher in the subsidized firms than in non subsidized ones, depending on the radius of the matching algorithm; the number of employees is from 11 to 29 per cent points higher and fixed assets increases to 24 per cent points. These results are all statistical significant. Also for productivity, profitability and financial state the results are quite similar to the structured case: there are no significant differences between financed and non financed firms.

Some differences between the two procedures (structured vs. non structured) come out by analyzing the relation between treatment levels and response variables. Again, we restrict the analysis only for outcome variables with a significant average treatment effect. Table 6.18 reports the OLS estimates together with the graphical representation of the predicted values (Figure 6.5). The results show a significant quadratic relation between effects and treatment levels only for employment growth. The profile is the same as before, with an increasing positive impact of the amount of incentives for low values of treatment level with respect to non treated firms. After reaching a maximum level, the effect of the subsidies on the outcomes is decreasing. For turnover and fixed asset variables the regression returns not significant values for the coefficients. In some sense, the structured case is able to better detect the heterogeneity of the effects with respect to different doses; this can be caused by a more efficient estimation of the selection process by exploiting more information. This finding is also confirmed by looking at the goodness of fit values: the structured case returns higher values for the adjusted R^2 (see Table 6.10 and Table 6.18).

As regards the quantile regression estimates (Table 6.19 and 6.20) the comparison of the structured with the non structured form of the selection process reveals no

	Employment			Turnover		Fixed Assets
	radius fix	radius mean	radius std	radius mean	radius std	radius mean
quota	0.931	1.822*	1.592*	-0.284	-0.221	0.942
quota ²	-0.016	-0.019*	-0.018*	0.001	-0.000	-0.006
const	21.72	-30.00	-16.91	18.62	20.99	-9.53
Obs.	66	66	66	66	66	65
Ad. R^2	0.131	0.060	0.051	-0.020	-0.015	-0.016
Prob. F	0.005	0.053	0.071	0.699	0.594	0.605

*p<0.05

Table 6.18. OLS estimates: impact of the amount of subsidies on the treatment level (non structured case)

differences between the two procedures in the coefficient estimates: again, the results confirm a significant quadratic relation between average effects and treatment levels for low percentile values of the dependent variable, especially for employment. However, the share of the impact variance explained by differences in the subsidies is higher in the case of a more structured form for the selection process (see Pseudo R^2 values in Table 6.11, 6.12 and Table 6.19, 6.20).

Finally figures 6.6 and 6.7 reports the results of the non parametric mean regressor estimator.

The graphical analysis confirm heterogeneity of the treatment effect with respect to different levels of treatment for the employment outcome variable. The tendency is quite similar to the structured case, with an increasing relation for low values of subsidies. Again, this kind of relation is robust to the radius size adopted in the matching procedure. Again, the value on which this maximum refers is between 50 and 60% of the share of subsidies on the investment. For turnover and fixed asset outcomes the relation between average effects and treatment level seems to be not significant, as confirmed by applying the parametric approaches above. Again, the structured case, with a more efficient estimation of the selection process, is able to better detect a significative relation and to better predict the heterogeneity of the impact with respect to the amount of granted subsidy.

Finally, independently from the chosen form for the selection process (structured or non structured) it is worth to keep in mind that in the analysis of the relation between impacts and treatment level, average causal effects are computed by comparing treated

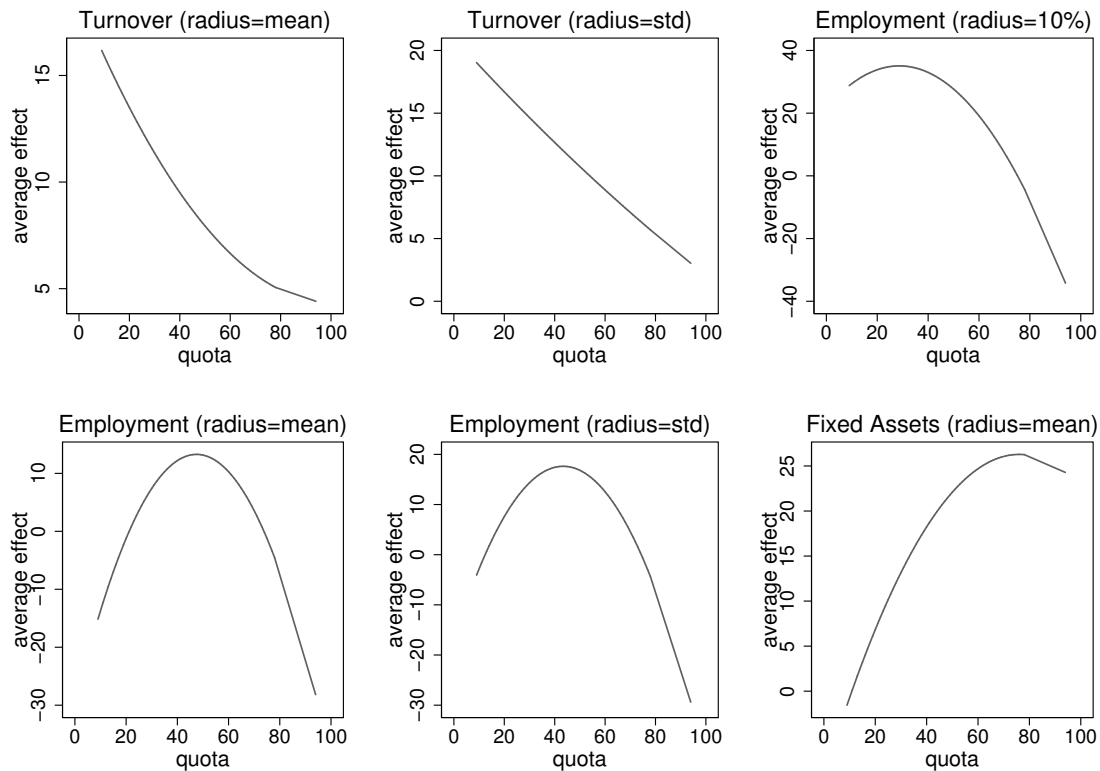


Figure 6.5. Predicted OLS estimates for treatment impact on the amount of subsidy (non structured case)

against non treated units. That is, our method is able to estimate the causal effect due to the participation ruling out the differences between these two groups. This means that, in general, there could be some differences among treated units (at different levels). That is, participants might be different not only with respect to the level of treatment. Then, in order to evaluate the impact of the amount of granted subsidy on treatment effects one should be careful in the interpretation of this relation. In the case of the Law 488, the proposed methods can explain heterogeneity but cannot suggest that, in the lower (higher) part of the curve, an increase in the amount of subsidy can increase (decrease) the impact. In fact, the characteristics of firms at different level of treatment can be different, and differences in the level of treatment can be imputed to this heterogeneity. In particular, this is true in this application: by construction the amount of granted subsidies is determined by the allocation procedure of Law 488,

		Employment					
		radius fix	Pseudo R^2	radius mean	Pseudo R^2	radius std	Pseudo R^2
q5	quota	8.092*	0.223	6.838*	0.152	7.850*	0.158
	quota ²	-0.093*		-0.077*		-0.088*	
	const	-166.9*		-154.1*		-171.0*	
q25	quota	3.260*	0.205	2.718*	0.133	2.156	0.132
	quota ²	-0.042*		-0.030*		-0.026	
	const	-39.8		-58.3*		-36.8	
q50	quota	0.400	0.087	1.969	0.079	1.521	0.067
	quota ²	-0.011		-0.019		-0.016	
	const	32.7		-39.1		-20.9	
q75	quota	-0.453	0.047	0.471	0.015	0.530	0.004
	quota ²	0.001		-0.004		-0.006	
	const	63.4		9.0		15.9	
q95	quota	1.198	0.068	1.857	0.035	1.628	0.027
	quota ²	-0.020		-0.018		-0.018	
	const	69.2		11.0		27.4	

*p<0.05, bootstrapped standard errors

Table 6.19. Quantile estimates: Employment (non structured case)

that imposes some constraints as function of geographic area and size of firms. Only if we use the same sample (the same firms' mix) at the different level of treatment this comparison is meaningful. We left this analysis for future research.

		Turnover				Fixed assets	
		radius mean	Pseudo R^2	radius std	Pseudo R^2	radius mean	Pseudo R^2
q5	quota	4.315*	0.219	4.656*	0.245	9.673*	0.226
	quota ²	-0.049*		-0.053*		-0.100*	
	const	-101.6*		-105.8*		-252.8*	
q25	quota	1.903	0.046	1.999	0.048	5.833*	0.088
	quota ²	-0.023*		-0.024*		-0.059	
	const	-40.1		-38.3		-145.7*	
q50	quota	0.692	0.019	0.684	0.026	4.104	0.029
	quota ²	-0.008		-0.008		-0.035	
	const	-7.0		-2.9		-88.8	
q75	quota	-0.816	0.017	-0.593	0.009	-0.993	0.070
	quota ²	0.010		0.007		0.019	
	const	32.4		32.3		46.2	
q95	quota	-6.839*	0.343	-6.778*	0.344	-5.372	0.097
	quota ²	0.072*		0.072*		0.039	
	const	200.9*		203.6*		272.8	

*p<0.05, bootstrapped standard errors

Table 6.20. Quantile estimates: Turnover and Fixed assets (non structured case)

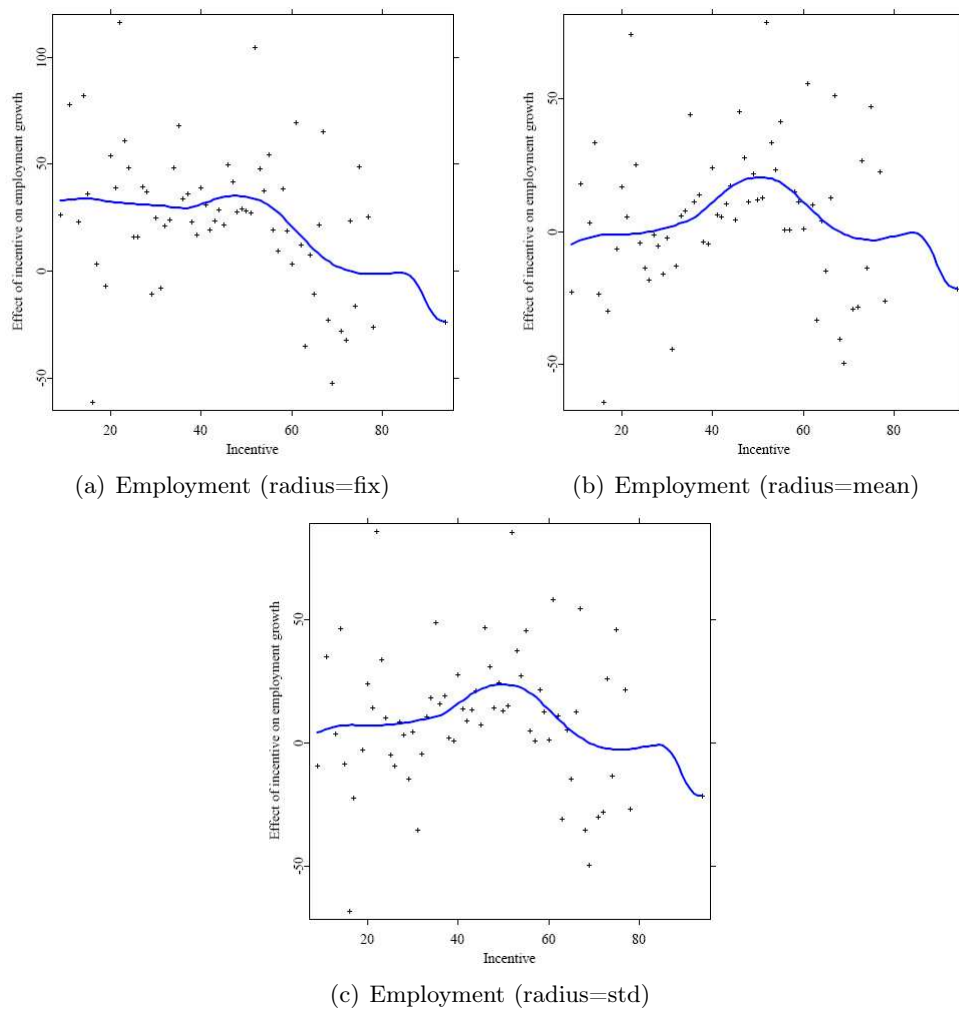


Figure 6.6. Kernel estimates for treatment impact on the amount of subsidy (non structured case)

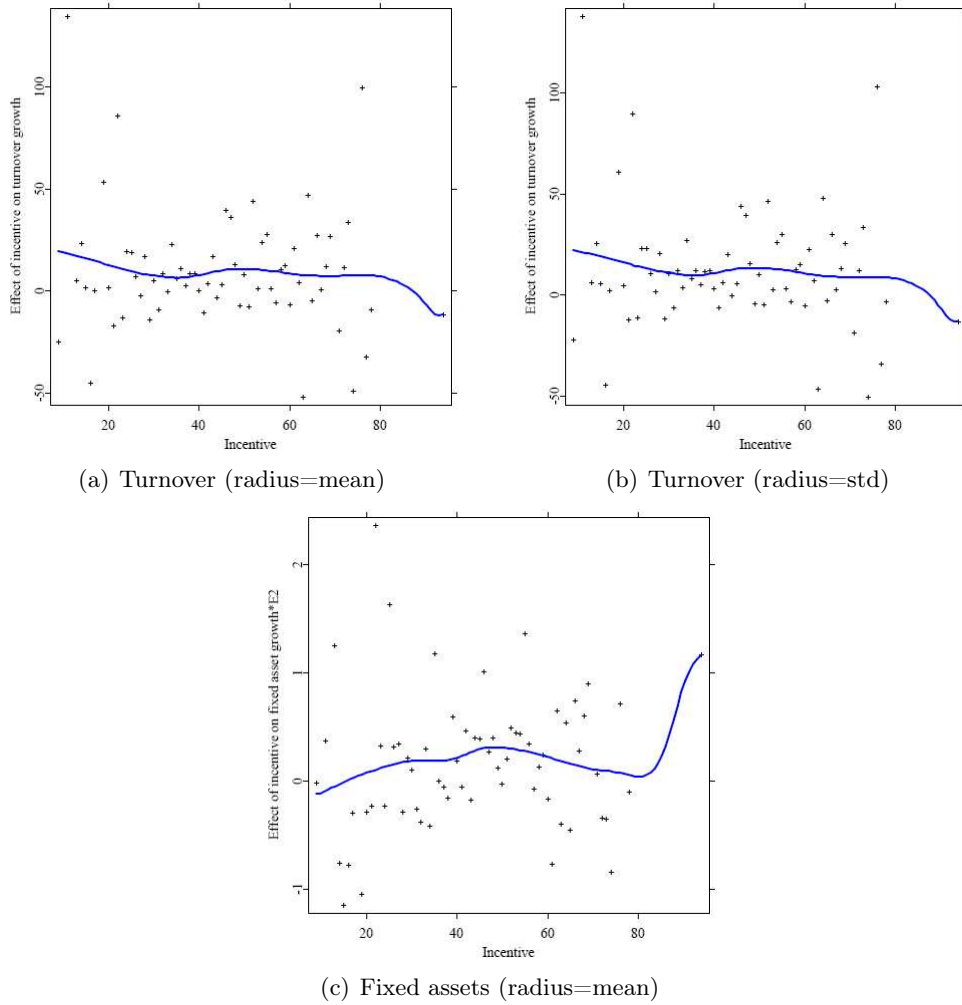


Figure 6.7. Kernel estimates for treatment impact on the amount of subsidy (non structured case)

Chapter 7

Conclusions

The ambitious aim of the thesis is to develop a matching estimator approach to evaluate causal effects of a policy intervention on some outcome variables in a continuous treatment framework. Recently matching estimators has been included in the new frontier tools for solving causal inference problems, such as program evaluation.

In order to solve an economic problem, that is the evaluation of the impacts of the Law 488/92 in Italy, we develop a novel double matching estimator, easy to apply and computationally not heavy. The proposed method allows to estimate average treatment effects at different level of treatment and subsequently to explore the impact of differences in treatment dose on policy outcome. Our results basically support the conclusions derived from methods based on the binary treatment (Pellegrini and Centra (2006) and Bernini *et al.* (2006)). Using the double matching method, the impact of L. 488 on subsidized firms is positive and statistically significant: the turnover increases from 9 to 12 points higher in the subsidized firms than in non subsidized ones, depending on matching algorithm; the number of employees is from 11 to 25 per cent points higher and fixed assets increases to 27 per cent points. The effects of the L. 488 are in line with the (explicit or less explicit) targets: the subsidized firms have invested more (in percent terms) than the non subsidized ones, achieving more turnover, more employment and more fixed assets. However, our methods show the strong heterogeneity of the treatment outcome with respect to different levels of treatment. The share of the impact variance explained by differences in the subsidies is about the 20% using the double matching method. We find that higher the level of incentive, higher the policy effect until a certain point, from which the marginal impact is decreasing.

Several economic policies use continuous policy variables. Therefore, this method

can have a wide application field. With respect to the methods proposed in literature in a continuous treatment setting, the two step matching method we introduce seems in some sense superior, because it offers more information: the impact to all different treatment level of the treated firms can be derived. Moreover, it evaluates the effects at each level of the subsidies by comparing treated versus non treated units. Furthermore, with the double matching procedure the selection process is estimated in a more efficient way, by splitting it into two components: the participation decision and the treatment level assignment. This “structured” form for the selection process is able to better detect the heterogeneity of the effects on outcome variables with respect to different treatment doses, rather than a case of a single selection process specification.

However, the method can explain heterogeneity but cannot suggest that, in the lower (higher) part of the curve, an increase in the amount of subsidy can increase (decrease) the impact. In fact, the characteristics of firms at different level of treatment can be different, and dissimilarities in the level of treatment can be imputed to this heterogeneity. Only if we use the same sample (the same firms’ mix) at the different level of treatment this comparison is meaningful. We left this analysis for future research. Moreover, as ulterior further development, we would like to indicate a robustness analysis on the matching algorithm in terms of distance measure among units.

Bibliography

- Abadie, A. and Imbens, G. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, **74**(1), 235–267. available at <http://ideas.repec.org/a/ecm/emetrp/v74y2006i1p235-267.html>.
- Angrist, J. (1990). Lifetime earnings and the vietnam era draft lottery: Evidence from social security administrative records. *American Economic Review*, **Vol. 80**(n. 3), pp. 313–36. available at <http://ideas.repec.org/a/aea/aecrev/v80y1990i3p313-36.html>.
- Angrist, J. and Krueger, A. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, **Vol. 106**(n. 4), pp. 979–1014. available at <http://ideas.repec.org/a/tpr/qjecon/v106y1991i4p979-1014.html>.
- Athey, S. and Imbens, G. (2006). Identification and inference in nonlinear difference-in-difference models. *Econometrica*, **vol. 74**(n.2), pp:431–497.
- Barca, F. and Pellegrini, G. (2002). *Policy for Territorial Competitiveness in Europe: Notes on the 2000-2006 Plan for the Italian Mezzogiorno*. Real Effects of Regional Integration in the European Union and the Mercosur: Inter-continental views on intra-continental experiences. Buenos Aires.
- Barnow, B., Cain, G., and Goldberger, A. (1980). *Issue in the Analysis of Selectivity Bias, in Evaluation Studies*, volume Vol. 5. E. Stromsdorfer and G. Farkas, san francisco: sage edition.
- Bartik, T. and Bingham, R. (1995). Can economic development programs be evaluated. *W.E. Upjohn Institute for Employment Research*.
- Battistin, E. and Rettore, E. (2003). Another look at the regression discontinuity design. CeMMAP working papers CWP01/03, Centre for Microdata Methods and Practice,

- Institute for Fiscal Studies. available at <http://ideas.repec.org/p/ifs/cemmap/01-03.html>.
- Behrman, J., Cheng, Y., and Todd, P. (2004). Evaluating preschool programs when length of exposure to the program varies: a non parametric approach. *Reviews of Economics and Statistics*, **Vol. 86**(N. 1), pp 108–132.
- Bell, B., Blundell, R., and Van Reenen, J. (1999). Getting the unemployed back to work: an evaluation of the new deal proposals. *International Tax and Public Finance*, **Vol. 6**, pp. 339–360.
- Bergstrom, F. (1998). Capital subsidies and the performance of firms. Working Paper No. 285, SSE/EFI Series in Economics and Finance, Department of Economics, University of Stockholm.
- Bernini, C., Centra, M., and Pellegrini, G. (2006). Growth and efficiency in subsidized firms. *Mimeo*.
- Blundell, R. and Costa Dias, M. (2002). Alternative approaches to evaluation in empirical microeconomics. CeMMAP working papers CWP10/02, Centre for Microdata Methods and Practice, Institute for Fiscal Studies. available at <http://ideas.repec.org/p/ifs/cemmap/10-02.html>.
- Blundell, R., Costa-Dias, M., Meghir, C., and Van Reenen, J. (2001). Evaluating the employment impact of a mandatory job search assistance program. IFS Working Papers W01/20, Institute for Fiscal Studies. available at <http://ideas.repec.org/p/ifs/ifsewp/01-20.html>.
- Boarnet, M. and Bogart, W. (1996). Enterprise zone and employment. what lessons can be learned? *ICER Torino Working Paper Series 98*.
- Bondonio, D. (2000). Statistical methods to evaluate geographically-target economic development programs. *Statistica Applicata*, **Vol. 12**(n. 2), pp. 177–204.
- Bondonio, D. (2004). The employment impact of business investment incentives in declining areas: an evaluation of the eu objective 2 area programs. *Università del Piemonte Orientale*.
- Bound, J., Jaeger, D., and Baker, R. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak. *Journal of the American Statistical Association*, **Vol. 90**(n. 430), pp. 443–450.

- Bronzini, R. and de Blasio, G. (2006). Evaluating the impact of investment incentives: The case of the italian law 488. *Journal of Urban Economics*, **Vol. 6**(n. 2), pp.
- Brown M. A., Curlee, R. T. and Elliott, S. R. (1995). Evaluating technology innovation programs: The use of comparison groups to identify impacts. *Research Policy*, **24**, pp. 669–684.
- Campbell, D. and Stanley, J. (1963). *Experimental and Quasi-Experimental Designs*. Chicago: Rand McNally.
- Card, D. and Robins, P. (1998). Do financial incentives encourage welfare recipients to work? *Research in Labour Economics*, **Vol. 17**, pp. 1–56.
- Carlucci, C. and Pellegrini, G. (2003). Gli effetti della legge 488/92: una valutazione dell'impatto occupazionale sulle imprese agevolate. *Rivista italiana degli Economisti*, **Vol. 8**(n. 2), pp. 267–286.
- Carlucci, C. and Pellegrini, G. (2005). Nonparametric analysis of the effects on employment of public subsidies to capital accumulation: the case of law 488/92 in Italy. presented at the Congress AIEL 2004, Modena.
- Chiri, S. and Pellegrini, G. (1995). Gli aiuti alle imprese nelle aree depresse. *Rivista economica del Mezzogiorno*, **n. 3**.
- Daly, M., Gorman, I., Lenjosek, G., MacNevin, A., and Phiriya-preunt, W. (1993). The impact of regional investment incentives on employment and productivity. *Regional Science and Urban Economics*, (23), pp. 559–575.
- Dehejia, R. (2005). Practical propensity score matching: a reply to Smith and Todd. *Journal of Econometrics*, **vol. 125**(issues 1-2), pp. 355–364.
- Dehejia, R. and Wahba, S. (1998a). Causal effects in non-experimental studies: Re-evaluating the evaluation of training programs. NBER Working Papers 6586, National Bureau of Economic Research, Inc. available at <http://ideas.repec.org/p/nbr/nberwo/6586.html>.
- Dehejia, R. and Wahba, S. (1998b). Propensity score matching methods for non-experimental causal studies. Technical report.
- Dowall, D. (1996). An evaluation of California's enterprise zone programs. *Economic Development Quarterly*, **Vol. 10**(n. 4), pp. 352–368.

- Driehuis, W. and van den Noord, P. (1998). The effects of investment subsidies on employment. *Economic Modelling*, **5**(1), pp. 32–40.
- Duflo, E. (2001). Schooling and labor market consequences of school construction in indonesia: Evidence from an unusual policy experiment. *American Economic Review*, **91**(4), 795–813. available at <http://ideas.repec.org/a/aea/aecrev/v91y2001i4p795-813.html>.
- Dupont, V. and Martin, P. (2003). Subsidies to poor regions and inequalities: Some unpleasant arithmetic. Technical Report No. 4107, CEPR Discussion Paper.
- Eissa, N. and Liebman, J. (1995). Labor supply response to the earned income tax credit. NBER Working Papers 5158, National Bureau of Economic Research, Inc. available at <http://ideas.repec.org/p/nbr/nberwo/5158.html>.
- Faini, R. and Schiantarelli, F. (1987). Incentives and investment decisions: the effectiveness of regional policy. *Oxford Economic Papers*, **39**, 516–533.
- Fisher, R. (1951). *The design of Experiments*. Edinburgh: Oliver and Boyd, 6th ed. edition.
- Florens, J., Heckman, J., Meghir, C., and Vytlacil, E. (????). Instrumental variable, local instrumental variables and control functions. *CeMMAP working papers CWP15/02, Centre for Microdata Methods and Practice, Institute for Fiscal Studies*.
- Flores, C. A. (2004). Estimation of dose response functions and optimal doses with continuous treatment. Working paper, University of Miami.
- Gabe, T. and Kraybill, D. (2002). The effects of state economic development incentives on employment growth of establishments. *Journal of Regional Science*, **42**, pp. 703–730.
- Garibaldi, P., Giavazzi, F., Ichino, A., and Rettore, E. (2007). College cost and time to complete a degree: Evidence from tuition discontinuities. working paper.
- Greenberg, D. and Shroder, M. (1997). *Digest of Social Experiments*. Urban Institute Press.
- Hahn, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica*, **66**(2), 315–332. available at <http://ideas.repec.org/a/ecm/emetrp/v66y1998i2p315-332.html>.

- Hahn, J., Todd, P., and Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, **69**(1), 201–09. available at <http://ideas.repec.org/a/ecm/emetrp/v69y2001i1p201-09.html>.
- Harris, R. (1991). The employment creation effects of factor subsidies: Some estimates for northern ireland manufacturing, 1955-83. *Journal of Regional Science*, **31**, pp. 49–64.
- Harris, R. and Trainor, M. (2005). Capital subsidies and their impact on total factor productivity: Firm-level evidence from northern ireland. *Journal of Regional Science*, **45**(1), pp. 49–74.
- Hausman, J. and Wise, D. (1985). *Social experimentation*. National Bureau of Economic Research.
- Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica*, **n.. 47**(pp. 153-161).
- Heckman, J. (1996). Randomization as an instrumental variable. *The Review of Economics and Statistics*, **Vol. 78**(N. 2), pp. 336–341.
- Heckman, J. and Robb, R. (1985). *Alternative methods for evaluating the impact of interventions*. Longitudinal Analysis of Labour Market Program.
- Heckman, J., Smith, J., and Clements, N. (1997a). Making the most out of programme evaluation and social experiment: accounting for heterogeneity in programme impacts. *Review of Economic Studies*, **Vol. 64**(n. 4), pp. 487–535.
- Heckman, J., Ichimura, H., and Todd, P. (1997b). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Review of Economic Studies*, **64**(4), 605–54. available at <http://ideas.repec.org/a/bla/restud/v64y1997i4p605-54.html>.
- Heckman, J., Ichimura, H., Smith, J., and Todd, P. (1998a). Characterizing selection bias using experimental data. NBER Working Papers 6699, National Bureau of Economic Research, Inc. available at <http://ideas.repec.org/p/nbr/nberwo/6699.html>.
- Heckman, J., Ichimura, H., and Todd, P. (1998b). Matching as an econometric evaluation estimator. *Review of Economic Studies*, **65**(2), 261–94. available at <http://ideas.repec.org/a/bla/restud/v65y1998i2p261-94.html>.

- Heckman, J., LaLonde, R., and Smith, J. (1999). The economics and econometrics of active labour market programs. *Handbook of Economics*, **3**.
- Hirano, K. and Imbens, G. (2004). The propensity score with continuous treatment. *Draft of a chapter for Missing data and bayesian methods in practise: contributions by Donald Rubin's Statistical family*. forthcoming from Wiley.
- Ichino, A. (2002). The problem of causality in the analysis of educational choices and labour market outcomes. *Lectures Notes*. <http://www.iue.it/Personal/Ichino/>.
- Ichino, A. and Becker, S. (2002). Estimation of average treatment effects based on propensity scores. *The Stata Journal*, **2**(4), 358–377.
- Ichino, A., Mealli, F., and Nannicini, T. (2003). Il lavoro interinale in italia: trappola del precariato o trampolino verso un impegno stabile? *Rapporto di Ricerca IUE*.
- Imai, K. and van Dyk, D. (2004). Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association*, **Vol. 99**(No. 467), pp. 854–866.
- Imbens, G. (1999). The role of the propensity score in estimating dose-response functions. NBER Technical Working Papers 0237, National Bureau of Economic Research, Inc. available at <http://ideas.repec.org/p/nbr/nberte/0237.html>.
- Imbens, G. (2003). Nonparametric estimation of average treatment effects under exogeneity: a review. Technical working paper 294, NATIONAL BUREAU OF ECONOMIC RESEARCH.
- Imbens, G. and Angrist, J. (1994). Identification and estimation of local average treatment effects. *Econometrica*, **Vol. 62**(n. 2), pp. 467–75. available at <http://ideas.repec.org/a/ecm/emetrp/v62y1994i2p467-75.html>.
- Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, **46**(33 50).
- LaLonde, R. (1986). Evaluating the econometric evaluations of training programs with experimental data. *American Econometric Review*, **Vol. 76**, pp.604–620.
- Lechner, M. (1999). Earnings and employment effects of continuous off-the-job training in east germany after unification. *Journal of Business and Economic Statistics*, **Vol. 17**(1), pp. 74–90.

- Lee, J. (1996). Government intervention and productivity growth. *Journal of Economic Growth*, **1**, pp. 391–414.
- Leuven, E. and Sianesi, B. (2003). Psmatch2: Stata module to perform full mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. Statistical Software Components, Boston College Department of Economics. available at <http://ideas.repec.org/c/boc/bocode/s432001.html>.
- Levitt, S. (1997). Using electoral cycles in police hiring to estimate the effect of police on crime. *American Economic Review*, **Vol. 87**(n. 3), pp. 270–90. available at <http://ideas.repec.org/a/aea/aecrev/v87y1997i3p270-90.html>.
- Lu, B., Zanutto, E., Hornik, R., and Rosenbaum, P. (2001). Matching with doses in an observational study of a media campaign against drug abuse. *Journal of the American Statistical Association*, **Vol. 96**(n. 456). Application and Case Studies.
- McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society*, (42), pp. 109–142.
- Midelfart-Knarvik, K. H. and Overman, G. (2002). Delocation and european integration: Is structural spending justified? *Economic Policy*, **v.17**(no. 35), pp. 321–359.
- Ministero dell'Industria, d. C. e. d. (2000). Relazione sulle leggi e sui provvedimenti di sostegno alle attività economiche e produttive. Technical report, Ministero dell'Industria, del Commercio e dell'Artigianato, Roma. various years.
- Nadayara, E. (1964). On estimating regression. *Theory and probability and its applications*, (vol. 9), pp:141–142.
- Orr, L. (1999). *Social Experiments: Evaluating Public Programs with Experimental Methods*. Sage Publication, thousand oaks, california edition.
- Orr, L., Bloom, H., Bell, S., Doolittle, F., and Lin, W. (1996). *Does Training for the Disadvantaged Work? Evidence from the National JTPA Study*. The Urban Institute Press.
- Parascandolo, P. and Pellegrini, G. (2001). Sistema d'asta ed efficienza nella valutazione del metodo di selezione delle imprese agevolate attraverso la legge 488/92. Atti del Convegno SIEP, Università di Pavia.
- Pellegrini, G. (1999). *L'efficacia degli aiuti alle imprese nel Mezzogiorno. Il vecchio e il nuovo intervento*. Il Mulino.

- Pellegrini, G. and Centra, M. (2006). Growth and efficiency in subsidized firms. Paper prepared for the Workshop "The Evaluation of Labour Market, Welfare and Firms Incentives Programmes", Istituto Veneto di Scienze, Lettere ed Arti - Venezia.
- Rooper, S. and Hewit-Dundas, N. (2001). Grant assistance and small firm development in northern ireland and the republic of ireland. *Scottish Journal of Political Economy*, **vol. 48**(n. 1), pp. 99–117.
- Rosenbaum, P. and Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, **70**(pp. 41-55).
- Royer, H. (2003). What all women (and some men) want to know: Does maternal age affect infant health? available at: <http://sitemaker.umich.edu/hroyer>).
- Rubin, D. (1973). Matching to remove bias in observational studies. *Biometrics*, **Vol. 29**, pp. 159–183.
- Rubin, D. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology*, (n. 66), pp. 666–701.
- Rubin, D. (1979). Using multivariate matched sampling and regression adjustment to control bias in observational studies. *Journal of the American Statistical Association*, **Vol. 74**(n. 366), pp. 318–328. available at <http://links.jstor.org/sici?sici=0162-1459R>
- Schalk, H. and Untied, G. (2002). Regional investment incentives in germany: Impacts on factor demand and growth. *Annals of Regional Sciences*, (34), pp. 173–195.
- Smith, J. and Todd, P. (2005a). Does matching overcome lalonde's critique of nonexperimental estimators? *Journal of Econometrics*, **vol. 125**(issues 1-2), pp. 305–353. available at <http://ideas.repec.org/p/uwo/hcuwoc/20035.html>.
- Smith, J. and Todd, P. (2005b). Rejoinder. *Journal of Econometrics*, **vol. 125**(issues 1-2), pp. 365–375.
- Stock, J. and Watson, M. (2003). *Introduction to Econometrics*. Boston, MA ; London: Addison Wesley.
- Trochim, W. (1984). *Research Design for Program Evaluation: the Regression-Discontinuity Approach*. Beverly Hills: Sage Publication, CA.

- van der Klaauw, W. (2002). Estimating the effect of financial aid offers on college enrollment: A regression-discontinuity approach. *International Economic Review*, **vol. 43**(n. 4), pp. 1249–1287. available at <http://ideas.repec.org/a/ier/iecrev/v43y2002i4p1249-1287.html>.
- Van Reenen, J. (1994). The creation and capture of rents: Wages and innovation in a panel of uk companies. CEPR Discussion Papers n. 1071, C.E.P.R. Discussion Papers. available at <http://ideas.repec.org/p/cpr/ceprdp/1071.html>.
- Willis, R. and Rosen, S. (1980). Education and self-selection. NBER Working Papers 0249, National Bureau of Economic Research, Inc. available at <http://ideas.repec.org/p/nbr/nberwo/0249.html>.
- Wren, C. and Waterson, M. (1991). The direct employment effect of financial assistance to industry. *Oxford Economic Paper*, **vol. 43**(n. 1), pp. 116–138.

Appendix A

Integration of datasets

Integration of L. 488 dataset with AIDA budgetary data

AIDA contains the budgets of firms whose turnover is more than 1 million euros, and therefore can not be representative of the Italian firms population. The budget data imputation procedure has produced an unavoidable reduction of the share of small firms in the sample, introducing a strong risk of selection bias in the composition of the treated and control groups. However, the analysis can be consistent if the selection bias is similar between the two groups: the under-representation of small firms in the sample could not affect the estimation of the policy impact if the variation in the small firms share is the same in the financed projects group and in the control group. The data suggest that the impact of the imputation procedure is basically the same between the two groups. The data suggest that the impact of the imputation procedure is basically the same between the two groups. The results show that the corrected sample, after the reduction in the data set due to identification of the eligible projects and elimination of anomalies, is equal to 34,9% of the original sample in subsidized firms and to 34,7% in non subsidized firms (table A.1¹).

The procedure of matching between administrative and budgetary data also reduces the sample: the matched sample used in the analysis is equal to 12,6% of the corrected sample for the subsidized firms, to 11,9% for non subsidized firms. The final dataset consists of 665 financed projects and 1.228 non financed projects.

Several careful checks of the matched data have also to be carried out in order

¹Source: elaboration of Pellegrini and Centra (2006) on L488 and AIDA data.

		Subsidized	Non Subsidized
Original projects dataset		15172	29586
1) Identification of eligible projects	Non ob. 1 regions	627	110
	Financed by other auctions	-	3726
	Presented in other auctions but non financed	-	3942
	Presented for more than one L. 488 auctions and financed almost once time	-	387
	Total	14545	21421
2) Discarded projects because out of the temporal lag	Financed but non concluded	7777	-
	Projects concluded (or scheduled to conclude) before the application year	200	320
	Projects started (or scheduled to start) before the year preceding the auction	980	4931
	Projects started (or scheduled to start) after 1999	293	5888
	Total	5295	10282
3) Budgetary data matching	Budgetary data in AIDA missing for the years used in the analysis	4630	9054
	Total	665	1228

Table A.1. From the original to the matched dataset

to measure the impact of the matching procedure. Our results critically depend on the absence of selection effect in the construction of the data set: the impact of the selection criteria and the missing data imputation procedure on financed projects and control group have to be deeply investigating. The analysis is presented in table A.2².

The reduction of the dataset due to the exclusion of AIDA missing firms, even if substantial in absolute value, has a slight impact on the regional distribution of firms. The variations systematically maintain the same sign for both the financed projects and the control group. Only in Puglia a higher reduction in financed projects with respect to the non financed ones is detected; this result is opposite to the one found in Basilicata.

²Source: elaboration of Pellegrini and Centra (2006) on L488 and AIDA data.

The impact on firm dimension distribution is analogous: the variations maintain the same sign for the three considered dimensional classes and there is no evidence of major differences.

The analysis of the distribution of projects according to firms economic activity shows a less favourable scenario. The matching process with AIDA reduces the incidence of firms operating in the extractive sector among financed projects group, while increases their share in the control group; the same happens for some sectors of mechanical industry. However, considering the whole sample of projects, this disparity on the sign of the variation in the share of economic activity among treated and control group regards only the 16% of the cases. The remaining 84%, has a distribution according to economic activity sector that does not vary in a significant way. The impact of the matching on the distributions of the two groups, subsidized and non subsidized, appears similar: the sign is the same in all distributions and the levels of differences do not show appreciable significant differences. These results support a high level of confidence on the data set representativeness and, consequently, on the robustness of the analysis. It is worth noting that activity sector and dimension are key variables in the evaluation procedure, since they are highly correlated with the social budgetary data: a high bias in these characteristics would cause a low reliability of results.

The matching with the AIDA dataset necessarily generates an asymmetry in the projects sample towards larger firms, for which the budgets are available and, indirectly, on the distribution of the indicators ranking the projects and on the level of the investment. As a consequence, a further step in the consistency analysis consists in evaluating the impact of AIDA integration on the selection indicators. The analysis shows a quite homogeneous impact on the indicators median value (less sensitive to outliers) which indicate an appreciable level of homogeneity between financed projects and control group.

		Not financed		Financed		Difference	
		Eligible projects (%)	Matched eligible projects (%)	Eligible projects (%)	Matched eligible projects (%)	Not financed	Financed
Geographical location	Molise	1,50%	2,52%	4,60%	4,23%	1,02%	-0,37%
	Campania	25,40%	34,04%	28,70%	34,34%	8,64%	5,64%
	Puglia	23,90%	23,94%	24,00%	22,40%	0,04%	-1,60%
	Basilicata	7,10%	6,68%	4,40%	5,81%	-0,42%	1,41%
	Calabria	21,40%	8,71%	15,30%	7,82%	-12,69%	-7,48%
	Sicilia	15,40%	17,02%	18,40%	18,81%	1,62%	0,41%
	Sardegna	5,30%	7,08%	4,60%	6,60%	1,78%	2,00%
	Total	100,00%	100,00%	100,00%	100,00%	0,00%	0,00%
Firm dimension	Large	2,70%	11,24%	4,70%	18,50%	8,54%	13,80%
	Medium	5,00%	21,99%	6,00%	27,37%	16,99%	21,37%
	Small	92,30%	66,78%	89,30%	54,14%	-25,52%	-35,16%
	Total	100,00%	100,00%	100,00%	100,00%	0,00%	0,00%
Economic activity sector	Mining and quarrying of energy producing materials	0,00%	0,00%	0,00%	0,15%	0,00%	0,15%
	Mining and quarrying, except of energy producing materials	2,20%	2,52%	1,70%	1,50%	0,32%	-0,20%
	Manufacture of food products, beverages and tobacco	19,40%	19,38%	9,20%	8,57%	-0,02%	-0,63%
	Manufacture of textiles and textile products	8,30%	5,05%	10,90%	8,42%	-3,25%	-2,48%
	Manufacture of leather and leather products	2,60%	2,77%	3,90%	4,96%	0,17%	1,06%
	Manufacture of wood and wood products	4,40%	1,95%	5,10%	2,41%	-2,45%	-2,69%
	Manufacture of pulp, paper and paper products; publishing and printing	6,20%	6,11%	5,40%	6,92%	-0,09%	1,52%
	Manufacture of coke, refined petroleum products and nuclear fuel	1,30%	3,50%	1,20%	3,61%	2,20%	2,41%
	Manufacture of chemicals, chemical products and man-made fibres	2,90%	4,48%	2,80%	6,17%	1,58%	3,37%
	Manufacture of rubber and plastic products	4,40%	6,03%	5,10%	9,32%	1,63%	4,22%
	Manufacture of other non-metallic mineral products	10,80%	11,24%	9,20%	10,23%	0,44%	1,03%
	Manufacture of basic metals and fabricated metal products	12,60%	13,52%	13,10%	13,23%	0,92%	0,13%
	Manufacture of machinery and equipment n.e.c.	3,70%	5,05%	5,20%	4,96%	1,35%	-0,24%
	Manufacture of electrical and optical equipment	3,90%	4,48%	5,90%	5,41%	0,58%	-0,49%
	Manufacture of transport equipment	2,30%	4,56%	2,20%	4,66%	2,26%	2,46%
	Manufacturing n.e.c.	6,70%	5,62%	7,20%	5,56%	-1,08%	-1,64%
	Unclassified service activities	8,30%	3,75%	11,90%	3,91%	-4,55%	-7,99%
	Total	100,00%	100,00%	100,00%	100,00%	0,00%	0,00%
Project typology	Modernization	6,90%	17,67%	3,90%	13,83%	10,77%	9,93%
	Extension	19,70%	42,18%	25,10%	53,08%	22,48%	27,98%
	Setting up	70,10%	34,28%	69,30%	28,27%	-35,82%	-41,03%
	Reactivation	0,20%	0,08%	0,20%	0,45%	-0,12%	0,25%
	Reconversion	0,10%	0,24%	0,10%	0,00%	0,14%	-0,10%
	Restructuring	2,90%	5,46%	1,40%	4,36%	2,56%	2,96%
	Relocation	0,10%	0,08%	0,00%	0,00%	-0,02%	0,00%
	Total	100,00%	100,00%	100,00%	100,00%	0,00%	0,00%

Table A.2. Impact of data matching process on eligible projects distribution

Appendix B

Outcome variable distributions

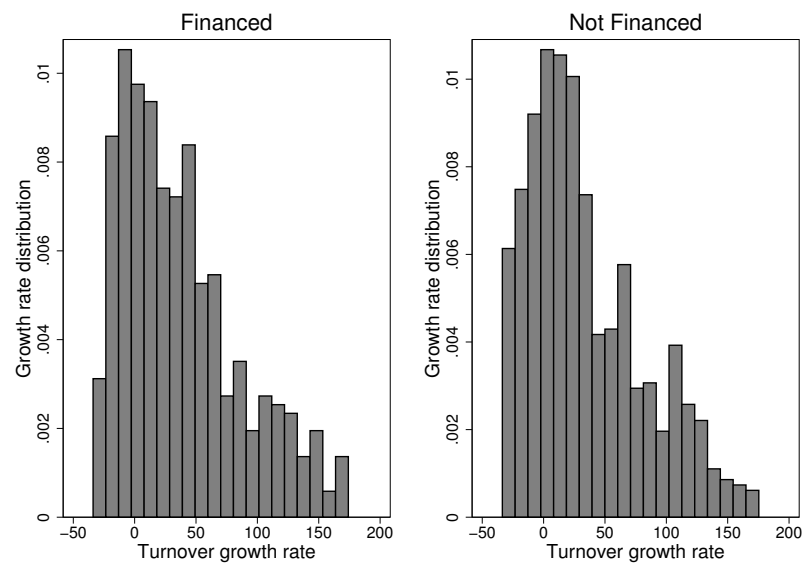


Figure B.1. Distribution of the outcome variable: turnover

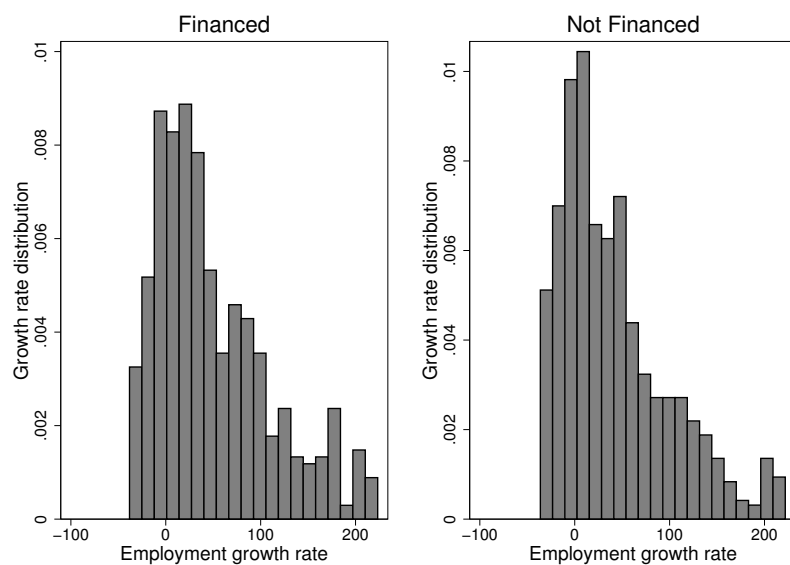


Figure B.2. Distribution of the outcome variable: employment

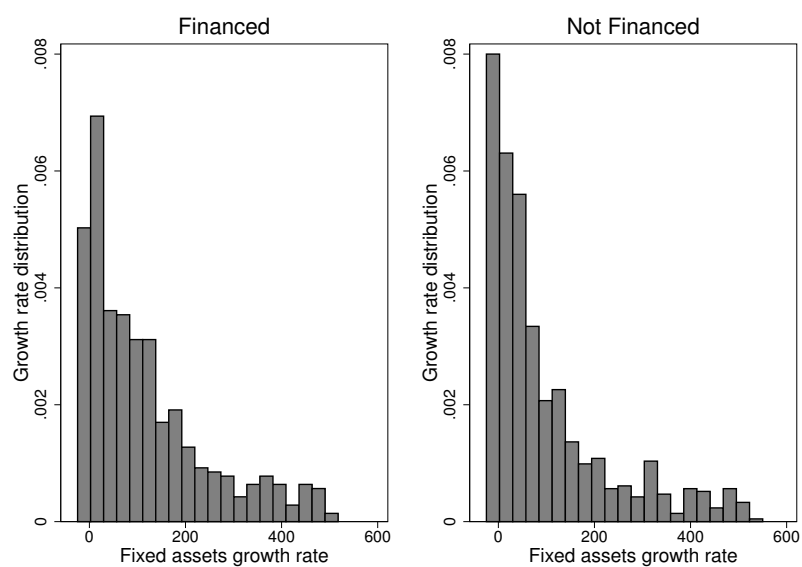


Figure B.3. Distribution of the outcome variable: fixed assets

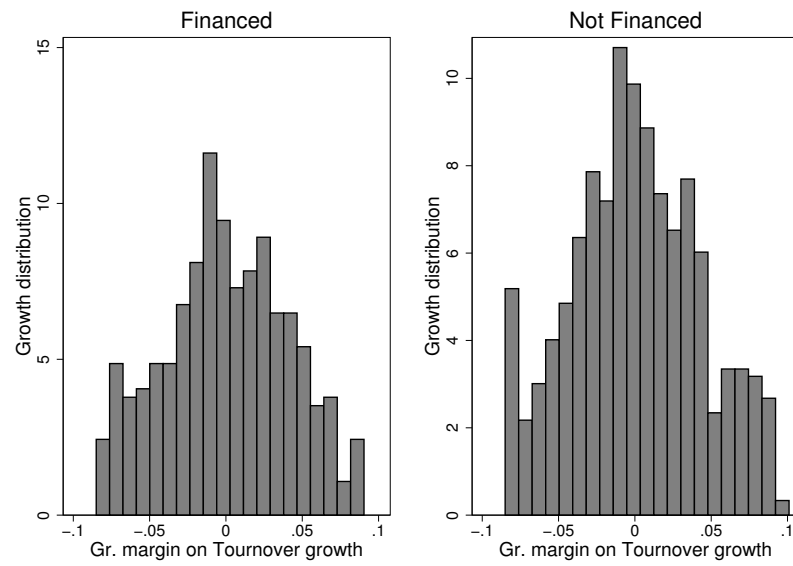


Figure B.4. Distribution of the outcome variable: gr. margin on turnover

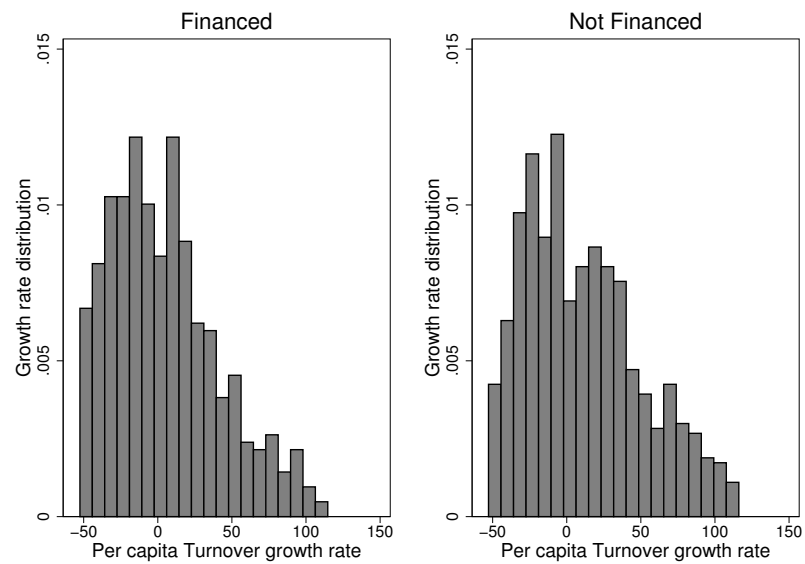


Figure B.5. Distribution of the outcome variable: per capita turnover

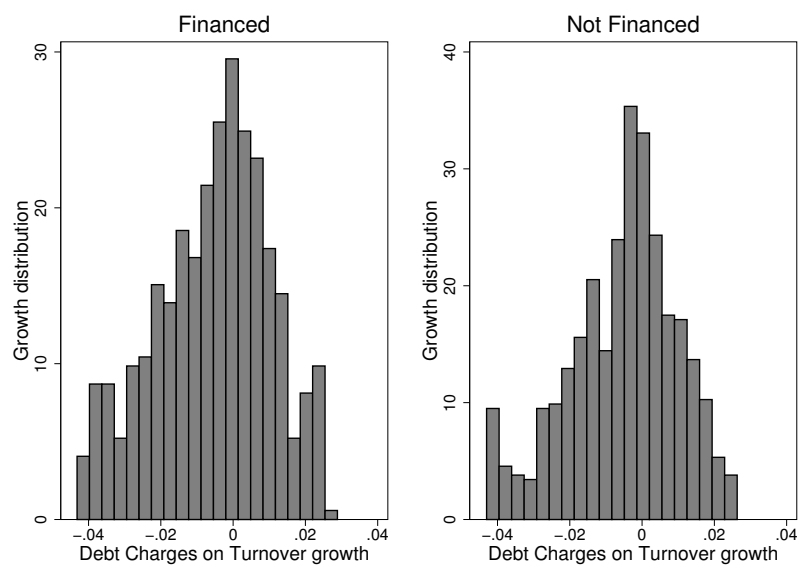


Figure B.6. Distribution of the outcome variable: debt charges on debt stock